

Our Ref.: CL001196

U.S. PATENT APPLICATION

Inventor(s): Gennady V. MERKULOV
Valentina DI FRANCESCO
Ellen M. BEASLEY

Invention: ISOLATED HUMAN RAS-LIKE PROTEINS, NUCLEIC ACID
MOLECULES ENCODING THESE HUMAN RAS-LIKE PROTEINS, AND
USES THEREOF

***CELERA GENOMICS CORPORATION.
45 WEST GUDE DR., C2-4#20
ROCKVILLE, MD 20850
(240) 453-3067
Fax (240)-453-3084***

SPECIFICATION

ISOLATED HUMAN RAS-LIKE PROTEINS, NUCLEIC ACID MOLECULES ENCODING THESE HUMAN RAS-LIKE PROTEINS, AND USES THEREOF

FIELD OF THE INVENTION

The present invention is in the field of Ras-like proteins that are related to the Rab subfamily, recombinant DNA molecules and protein production. The present invention specifically provides novel Ras-like protein polypeptides and proteins and nucleic acid molecules encoding such peptide and protein molecules, all of which are useful in the development of human therapeutics and diagnostic compositions and methods.

BACKGROUND OF THE INVENTION

Ras-like proteins, particularly members of the Rab subfamilies, are a major target for drug action and development. Accordingly, it is valuable to the field of pharmaceutical development to identify and characterize previously unknown members of this subfamily of Ras-like proteins. The present invention advances the state of the art by providing a previously unidentified human Ras-like proteins that have homology to members of the Rab subfamilies.

Ras protein

Ras proteins are small regulatory GTP-binding proteins, or small G proteins, which belong to the Ras protein superfamily. They are monomeric GTPases, but their GTPase activity is very slow (less than one GTP molecule per minute).

Ras proteins are key relays in the signal-transducing cascade induced by the binding of a ligand to specific receptors such as receptor tyrosine kinases (RTKs), since they trigger the MAP kinase cascade. The ligand can be a growth factor (epidermal growth factor (EGF), platelet-derived growth factor (PDGF), insulin, an interleukin (IL), granulocyte colony-stimulating factor (G-CSF), granulocyte/macrophage colony-stimulating factor (GM-CSF).

Ras proteins contain sequences highly conserved during evolution. Their tertiary structure includes ten loops connecting six strands of beta-sheet and five alpha helices.

In mammals, there are four Ras proteins, which are encoded by Ha-ras, N-ras, Ki-rasA and Ki-rasB genes. They are composed of about 170 residues and have a relative molecular mass of 21 kD. Ras proteins contain covalently attached modified lipids allowing these proteins to bind to the plasma membrane. Ha-Ras has a C-terminal farnesyl group, a C-terminal palmitoyl group and a N-terminal myristoyl group. In Ki-Ras(B), a C-terminal polylysine domain replaces the palmitoyl group.

Ras proteins alternate between an inactive form bound to GDP and an active form bound to GTP. Their activation results from reactions induced by a guanine nucleotide-exchange factor (GEF). Their inactivation results from reactions catalyzed by a GTPase-activating protein (GAP).

When a Ras protein is activated by a GEF such as a Sos protein, the N-terminal region of a serine/threonine kinase, called "Raf protein", can bind to Ras protein. The C-terminal region of the activated Raf thus formed binds to another protein, MEK, and phosphorylates it on both specific tyrosine and serine residues. Active MEK phosphorylates and activates, in turn, a MAP kinase (ERK1 or ERK2), which is also a serine/threonine kinase. This phosphorylation occurs on both specific tyrosine and threonine residues of MAP kinase.

MAP kinase phosphorylates many different proteins, especially nuclear transcription factors (TFs) that regulate expression of many genes during cell proliferation and differentiation.

Recent researches suggest that, in mammals, phosphatidyl inositol 3'-kinase (PI3-kinase) might be a target of Ras protein, instead of Raf protein. In certain mutations, the translation of ras genes may produce oncogenic Ras proteins.

Ras-like protein

Guanine nucleotide-binding proteins (GTP-binding proteins, or G proteins) participate in a wide range of regulatory functions including metabolism, growth, differentiation, signal transduction, cytoskeletal organization, and intracellular vesicle transport and secretion. These proteins control diverse sets of regulatory pathways in

response to hormones, growth factors, neuromodulators, or other signaling molecules. When these molecules bind to transmembrane receptors, signals are propagated to effector molecules by intracellular signal transducing proteins. Many of these signal-transducing proteins are members of the Ras superfamily.

The Ras superfamily is a class of low molecular weight (LMW) GTP-binding proteins that consist of 21-30 kDa polypeptides. These proteins regulate cell growth, cell cycle control, protein secretion, and intracellular vesicle interaction. In particular, the LMW GTP-binding proteins activate cellular proteins by transducing mitogenic signals involved in various cell functions in response to extracellular signals from receptors (Tavitian, A. (1995) C. R. Seances Soc. Biol. Fil. 189:7-12). During this process, the hydrolysis of GTP acts as an energy source as well as an on-off switch for the GTPase activity of the LMW GTP-binding proteins.

The Ras superfamily is comprised of five subfamilies: Ras, Rho, Ran, Rab, and ADP-ribosylation factor (ARF). Specifically, Ras genes are essential in the control of cell proliferation. Mutations in Ras genes have been associated with cancer. Rho proteins control signal transduction in the process of linking receptors of growth factors to actin polymerization that is necessary for cell division. Rab proteins control the translocation of vesicles to and from membranes for protein localization, protein processing, and secretion. Ran proteins are localized to the cell nucleus and play a key role in nuclear protein import, control of DNA synthesis, and cell-cycle progression. ARF and ARF-like proteins participate in a wide variety of cellular functions including vesicle trafficking, exocrine secretion, regulation of phospholipase activity, and endocytosis.

Despite their sequence variations, all five subfamilies of the Ras superfamily share conserved structural features. Four conserved sequence regions (motifs I-IV) have been studied in the LMW GTP-binding proteins. Motif I is the most variable but has the conserved sequence, GXXXXGK. The lysine residue is essential in interacting with the .beta.- and .gamma.-phosphates of GTP. Motif II, III, and IV contain highly conserved sequences of DTAGQ, NKXD, and EXSAX, respectively. Specifically, Motif II regulates the binding of gamma-phosphate of GTP; Motif III regulates the binding of GTP; and Motif IV regulates the guanine base of GTP. Most of the membrane-bound LMW GTP-binding proteins generally require a carboxy terminal isoprenyl group for membrane

association and biological activity. The isoprenyl group is added posttranslationally through recognition of a terminal cysteine residue alone or a terminal cysteine-aliphatic amino acid-aliphatic amino acid-any amino acid (CAAX) motif. Additional membrane-binding energy is often provided by either internal palmitoylation or a carboxy terminal cluster of basic amino acids. The LMW GTP-binding proteins also have a variable effector region, located between motifs I and II, which is characterized as the interaction site for guanine nucleotide exchange factors (GEFs) or GTPase-activating proteins (GAPs). GEFs induce the release of GDP from the active form of the G protein, whereas GAPs interact with the inactive form by stimulating the GTPase activity of the G protein.

The ARF subfamily has at least 15 distinct members encompassing both ARF and ARF-like proteins. ARF proteins identified to date exhibit high structural similarity and ADP-ribosylation enhancing activity. In contrast, several ARF-like proteins lack ADP-ribosylation enhancing activity and bind GTP differently. An example of ARF-like proteins is a rat protein, ARL184. ARL184 has been shown to have a molecular weight of 22 kDa and four functional GTP-binding sites (Icard-Liepkalns, C. et al. (1997) Eur. J. Biochem. 246: 388-393). ARL184 is active in both the cytosol and the Golgi apparatus and is closely associated with acetylcholine release, suggesting that ARL184 is a potential regulatory protein associated with Ca²⁺-dependent release of acetylcholine.

A number of Rho GTP-binding proteins have been identified in plasma membrane and cytoplasm. These include RhoA, B and C, and D, rhoG, rac 1 and 2, G25K-A and B, and TC10 (Hall, A. et al. (1993) Philos. Trans. R. Soc. Lond. (Biol.) 340:267-271). All Rho proteins have a CAAX motif that binds a prenyl group and either a palmitoylation site or a basic amino acid-rich region, suggesting their role in membrane-associated functions. In particular, RhoD is a protein that functions in early endosome motility and distribution by inducing rearrangement of actin cytoskeleton and cell surface (Murphy, C. et al. (1996) Nature 384:427-432). During cell adhesion, the Rho proteins are essential for triggering focal complex assembly and integrin-dependent signal transduction (Hotchin, N. A. and Hall, A. (1995) J. Cell Biol. 131:1857-1865).

The Ras subfamily proteins already indicated supra are essential in transducing signals from receptor tyrosine kinases (RTKs) to a series of serine/threonine kinases

which control cell growth and differentiation. Mutant Ras proteins, which bind but cannot hydrolyze GTP, are permanently activated and cause continuous cell proliferation or cancer. TC21, a Ras-like protein, is found to be highly expressed in a human teratocarcinoma cell line (Drivas, G. T. et al. (1990) *Mol. Cell. Biol.* 10: 1793-1798). Rin and Rit are characterized as membrane-binding, Ras-like proteins without the lipid-binding CAAX motif and carboxy terminal cysteine (Lee, C.-H. J. et al. (1996) *J. Neurosci.* 16: 6784-6794). Further, Rin is shown to localize in neurons and have calcium-dependant calmodulin-binding activity.

Rab proteins

The novel human protein, and encoding gene, provided by the present invention is related to the Rab family of Ras-like proteins and shows the highest degree of similarity to Rab1. Rab GTP-binding proteins are similar to YPT1/SEC4 in *Saccharomyces cerevisiae*, which are critical for transport along the exocytic route (Chavrier *et al.*, *Mol Cell Biol* 1990 Dec;10(12):6578-85). Different Rab proteins are presumed to control different steps in membrane traffic, leading to a high level of diversity and complexity within the Rab family (Chavrier *et al.*, *Mol Cell Biol* 1990 Dec;10(12):6578-85). The Rab1 gene maps in close vicinity to the 'wobbler' spinal muscular atrophy gene.

RAB proteins are important for regulating the targeting and fusion of membranous vesicles during organelle assembly and transport. RAB proteins undergo controlled exchange of GTP for GDP, and they hydrolyze GTP in a reaction that may regulate the timing and unidirectional nature of these assemblies. Generally, known RAB proteins terminate in sequences such as cys-X-cys (e.g., RAB3A), cys-cys (e.g., RAB1A), or a similar sequence, and generally all are geranylgeranylated.

The tethering factor p115 is a RAB1 effector that binds directly to activated RAB1. It is thought that RAB1-regulated assembly of functional effector-SNARE complexes serves as a conserved molecular mechanism for regulating recognition between different subcellular compartments such as endoplasmic reticulum and Golgi apparatus (Allan *et al.*, *Science* 289: 444-448, 2000).

GTPases play important roles in a wide variety of cell functions such as signal transduction, cytoskeletal organization, and membrane trafficking. Rab GTPases are

particularly important for regulating cellular membrane dynamics by modulating the activity of effector proteins that then regulate vesicle trafficking. The Rab8 GTPase plays important roles in Golgi to plasma membrane vesicle trafficking. Studies have suggested that Rab37 plays an important role in mast cell degranulation. Thus, novel human Rab GTPases may be valuable as potential therapeutic targets for the development of allergy treatments (Masuda *et al.*, *FEBS Lett* 2000 Mar 17;470). Rab15 may act, together with Rab3A, to regulate synaptic vesicle membrane flow within nerve terminals, thereby regulating neurotransmitter release. Rab15 and Rab3A are low molecular weight GTP-binding proteins. Rab proteins are generally comprised of four conserved structural domains necessary for GTP binding, as well as additional domains for membrane localization and effector protein interactions. Rab15 is expressed primarily in neural tissues such as the brain and is localized to synaptic vesicles (Elferink *et al.*, *J. Biol. Chem.* 267 (9), 5768-5775 (1992)).

For a further review of Rab1 and other Rab proteins, see Wedemeyer *et al.*, *Genomics* 32: 447-454, 1996 and Zahraoui *et al.*, *J. Biol. Chem.* 264: 12394-12401, 1989.

Due to their importance in human physiology, particularly in regulating membrane trafficking, novel human Rab proteins/genes, such as provided by the present invention, are valuable as potential targets for the development of therapeutics to treat a wide variety of diseases/disorders caused or influenced by defects in membrane trafficking. Furthermore, SNPs in Rab genes, such as provided by the present invention, are valuable markers for the diagnosis, prognosis, prevention, and/or treatment of such diseases/disorders.

Using the information provided by the present invention, reagents such as probes/primers for detecting the SNPs or the expression of the protein/gene provided herein may be readily developed and, if desired, incorporated into kit formats such as nucleic acid arrays, primer extension reactions coupled with mass spec detection (for SNP detection), or TaqMan PCR assays (Applied Biosystems, Foster City, CA).

The discovery of new human Ras-like proteins and the polynucleotides that encode them satisfies a need in the art by providing new compositions that are useful in

the diagnosis, prevention, and treatment of inflammation and disorders associated with cell proliferation and apoptosis.

SUMMARY OF THE INVENTION

The present invention is based in part on the identification of amino acid sequences of human Ras-like protein polypeptides and proteins that are related to the Ras-like protein subfamily, as well as allelic variants and other mammalian orthologs thereof. These unique peptide sequences, and nucleic acid sequences that encode these peptides, can be used as models for the development of human therapeutic targets, aid in the identification of therapeutic proteins, and serve as targets for the development of human therapeutic agents that modulate Ras-like protein activity in cells and tissues that express the Ras-like protein. Experimental data as provided in Figure 1 indicates expression in the adult and fetal brain, adrenal gland, brain neuroblastoma, skin melanotic melanoma, uterus leiomyosarcoma, and placenta.

DESCRIPTION OF THE FIGURE SHEETS

FIGURE 1 provides the nucleotide sequence of a cDNA molecule that encodes the Ras-like protein of the present invention. (SEQ ID NO:1) In addition, structure and functional information is provided, such as ATG start, stop and tissue distribution, where available, that allows one to readily determine specific uses of inventions based on this molecular sequence. Experimental data as provided in Figure 1 indicates expression in the adult and fetal brain, adrenal gland, brain neuroblastoma, skin melanotic melanoma, uterus leiomyosarcoma, and placenta.

FIGURE 2 provides the predicted amino acid sequence of the Ras-like protein of the present invention. (SEQ ID NO:2) In addition structure and functional information such as protein family, function, and modification sites is provided where available, allowing one to readily determine specific uses of inventions based on this molecular sequence.

FIGURE 3 provides genomic sequences that span the gene encoding the Ras-like protein of the present invention. (SEQ ID NO:3) In addition structure and functional

information, such as intron/exon structure, promoter location, etc., is provided where available, allowing one to readily determine specific uses of inventions based on this molecular sequence. As illustrated in Figure 3, SNPs were identified at 25 different nucleotide positions.

DETAILED DESCRIPTION OF THE INVENTION

General Description

The present invention is based on the sequencing of the human genome. During the sequencing and assembly of the human genome, analysis of the sequence information revealed previously unidentified fragments of the human genome that encode peptides that share structural and/or sequence homology to protein/peptide/domains identified and characterized within the art as being a Ras-like protein or part of a Ras-like protein and are related to the Rab subfamily. Utilizing these sequences, additional genomic sequences were assembled and transcript and/or cDNA sequences were isolated and characterized. Based on this analysis, the present invention provides amino acid sequences of human Ras-like protein polypeptides that are related to the Rab subfamily, nucleic acid sequences in the form of transcript sequences, cDNA sequences and/or genomic sequences that encode these Ras-like protein polypeptide, nucleic acid variation (allelic information), tissue distribution of expression, and information about the closest art known protein/peptide/domain that has structural or sequence homology to the Ras-like protein of the present invention.

In addition to being previously unknown, the peptides that are provided in the present invention are selected based on their ability to be used for the development of commercially important products and services. Specifically, the present peptides are selected based on homology and/or structural relatedness to known Ras-like proteins of the Rab subfamily and the expression pattern observed. Experimental data as provided in Figure 1 indicates expression in the adult and fetal brain, adrenal gland, brain neuroblastoma, skin melanotic melanoma, uterus leiomyosarcoma, and placenta. The art has clearly established the commercial importance of members of this family of proteins and proteins that have expression patterns similar to that of the present gene. Some of the

more specific features of the peptides of the present invention, and the uses thereof, are described herein, particularly in the Background of the Invention and in the annotation provided in the Figures, and/or are known within the art for each of the known Rab family or subfamily of Ras-like proteins.

Specific Embodiments

Peptide Molecules

The present invention provides nucleic acid sequences that encode protein molecules that have been identified as being members of the Ras-like protein family and are related to the Rab subfamily (protein sequences are provided in Figure 2, transcript/cDNA sequences are provided in Figure 1 and genomic sequences are provided in Figure 3). The peptide sequences provided in Figure 2, as well as the obvious variants described herein, particularly allelic variants as identified herein and using the information in Figure 3, will be referred herein as the Ras-like proteins or peptides of the present invention, Ras-like proteins or peptides, or peptides/proteins of the present invention.

The present invention provides isolated peptide and protein molecules that consist of, consist essentially of, or comprise the amino acid sequences of the Ras-like protein polypeptide disclosed in the Figure 2, (encoded by the nucleic acid molecule shown in Figure 1, transcript/cDNA or Figure 3, genomic sequence), as well as all obvious variants of these peptides that are within the art to make and use. Some of these variants are described in detail below.

As used herein, a peptide is said to be "isolated" or "purified" when it is substantially free of cellular material or free of chemical precursors or other chemicals. The peptides of the present invention can be purified to homogeneity or other degrees of purity. The level of purification will be based on the intended use. The critical feature is that the preparation allows for the desired function of the peptide, even if in the presence of considerable amounts of other components.

In some uses, "substantially free of cellular material" includes preparations of the peptide having less than about 30% (by dry weight) other proteins (i.e., contaminating

protein), less than about 20% other proteins, less than about 10% other proteins, or less than about 5% other proteins. When the peptide is recombinantly produced, it can also be substantially free of culture medium, i.e., culture medium represents less than about 20% of the volume of the protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of the peptide in which it is separated from chemical precursors or other chemicals that are involved in its synthesis. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of the Ras-like protein polypeptide having less than about 30% (by dry weight) chemical precursors or other chemicals, less than about 20% chemical precursors or other chemicals, less than about 10% chemical precursors or other chemicals, or less than about 5% chemical precursors or other chemicals.

The isolated Ras-like protein polypeptide can be purified from cells that naturally express it, purified from cells that have been altered to express it (recombinant), or synthesized using known protein synthesis methods. Experimental data as provided in Figure 1 indicates expression in the adult and fetal brain, adrenal gland, brain neuroblastoma, skin melanotic melanoma, uterus leiomyosarcoma, and placenta. For example, a nucleic acid molecule encoding the Ras-like protein polypeptide is cloned into an expression vector, the expression vector introduced into a host cell and the protein expressed in the host cell. The protein can then be isolated from the cells by an appropriate purification scheme using standard protein purification techniques. Many of these techniques are described in detail below.

Accordingly, the present invention provides proteins that consist of the amino acid sequences provided in Figure 2 (SEQ ID NO:2), for example, proteins encoded by the transcript/cDNA nucleic acid sequences shown in Figure 1 (SEQ ID NO:1) and the genomic sequences provided in Figure 3 (SEQ ID NO:3). The amino acid sequence of such a protein is provided in Figure 2. A protein consists of an amino acid sequence when the amino acid sequence is the final amino acid sequence of the protein.

The present invention further provides proteins that consist essentially of the amino acid sequences provided in Figure 2 (SEQ ID NO:2), for example, proteins encoded by the transcript/cDNA nucleic acid sequences shown in Figure 1 (SEQ ID NO:1) and the genomic

sequences provided in Figure 3 (SEQ ID NO:3). A protein consists essentially of an amino acid sequence when such an amino acid sequence is present with only a few additional amino acid residues, for example from about 1 to about 100 or so additional residues, typically from 1 to about 20 additional residues in the final protein.

The present invention further provides proteins that comprise the amino acid sequences provided in Figure 2 (SEQ ID NO:2), for example, proteins encoded by the transcript/cDNA nucleic acid sequences shown in Figure 1 (SEQ ID NO:1) and the genomic sequences provided in Figure 3 (SEQ ID NO:3). A protein comprises an amino acid sequence when the amino acid sequence is at least part of the final amino acid sequence of the protein. In such a fashion, the protein can be only the peptide or have additional amino acid molecules, such as amino acid residues (contiguous encoded sequence) that are naturally associated with it or heterologous amino acid residues/peptide sequences. Such a protein can have a few additional amino acid residues or can comprise several hundred or more additional amino acids. The preferred classes of proteins that are comprised of the Ras-like protein polypeptide of the present invention are the naturally occurring mature proteins. A brief description of how various types of these proteins can be made/isolated is provided below.

The Ras-like protein polypeptides of the present invention can be attached to heterologous sequences to form chimeric or fusion proteins. Such chimeric and fusion proteins comprise a Ras-like protein polypeptide operatively linked to a heterologous protein having an amino acid sequence not substantially homologous to the Ras-like protein polypeptide. "Operatively linked" indicates that the Ras-like protein polypeptide and the heterologous protein are fused in-frame. The heterologous protein can be fused to the N-terminus or C-terminus of the Ras-like protein polypeptide.

In some uses, the fusion protein does not affect the activity of the Ras-like protein polypeptide *per se*. For example, the fusion protein can include, but is not limited to, enzymatic fusion proteins, for example beta-galactosidase fusions, yeast two-hybrid GAL fusions, poly-His fusions, MYC-tagged, HI-tagged and Ig fusions. Such fusion proteins, particularly poly-His fusions, can facilitate the purification of recombinant Ras-like protein polypeptide. In certain host cells (e.g., mammalian host cells), expression and/or secretion of a protein can be increased by using a heterologous signal sequence.

A chimeric or fusion protein can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different protein sequences are ligated together in-frame in accordance with conventional techniques. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and re-amplified to generate a chimeric gene sequence (see Ausubel *et al.*, *Current Protocols in Molecular Biology*, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST protein). A Ras-like protein polypeptide-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the Ras-like protein polypeptide.

As mentioned above, the present invention also provides and enables obvious variants of the amino acid sequence of the peptides of the present invention, such as naturally occurring mature forms of the peptide, allelic/sequence variants of the peptides, non-naturally occurring recombinantly derived variants of the peptides, and orthologs and paralogs of the peptides. Such variants can readily be generated using art know techniques in the fields of recombinant nucleic acid technology and protein biochemistry. It is understood, however, that variants exclude any amino acid sequences disclosed prior to the invention.

Such variants can readily be identified/made using molecular techniques and the sequence information disclosed herein. Further, such variants can readily be distinguished from other peptides based on sequence and/or structural homology to the Ras-like protein polypeptides of the present invention. The degree of homology/identity present will be based primarily on whether the peptide is a functional variant or non-functional variant, the amount of divergence present in the paralog family, and the evolutionary distance between the orthologs.

To determine the percent identity of two amino acid sequences or two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for

comparison purposes). In a preferred embodiment, the length of a reference sequence aligned for comparison purposes is at least 30%, 40%, 50%, 60%, 70%, 80%, or 90% or more of the length of the reference sequence. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

The comparison of sequences and determination of percent identity and similarity between two sequences can be accomplished using a mathematical algorithm. (*Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part 1*, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (*J. Mol. Biol.* (48):444-453 (1970)) algorithm which has been incorporated into the GAP program in the GCG software package (available at <http://www.gcg.com>), using either a Blossom 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (Devereux, J., *et al.*, *Nucleic Acids Res.* 12(1):387 (1984)) (available at <http://www.gcg.com>), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. In another embodiment, the percent identity between two amino acid or nucleotide sequences is determined using the algorithm of E. Meyers and W. Miller (*CABIOS*, 4:11-

17 (1989)) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

The nucleic acid and protein sequences of the present invention can further be used as a "query sequence" to perform a search against sequence databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, et al. (*J. Mol. Biol.* 215:403-10 (1990)). BLAST nucleotide searches can be performed with the NBLAST program, score = 100, word length = 12 to obtain nucleotide sequences homologous to the nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, word length = 3, to obtain amino acid sequences homologous to the proteins of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al. (*Nucleic Acids Res.* 25(17):3389-3402 (1997)). When utilizing BLAST and gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. See <http://www.ncbi.nlm.nih.gov>.

Full-length pre-processed forms, as well as mature processed forms, of proteins that comprise one of the peptides of the present invention can readily be identified as having complete sequence identity to one of the Ras-like protein polypeptides of the present invention as well as being encoded by the same genetic locus as the Ras-like protein polypeptide provided herein. The gene encoding the novel Ras-like protein of the present invention is located on a genome component that has been mapped to human chromosome 2 (as indicated in Figure 3), which is supported by multiple lines of evidence, such as STS and BAC map data.

Allelic variants of a Ras-like protein polypeptide can readily be identified as being a human protein having a high degree (significant) of sequence homology/identity to at least a portion of the Ras-like protein polypeptide as well as being encoded by the same genetic locus as the Ras-like protein polypeptide provided herein. Genetic locus can readily be determined based on the genomic information provided in Figure 3, such as the genomic sequence mapped to the reference human. The gene encoding the novel Ras-like protein of the present invention is located on a genome component that has been mapped to human chromosome 2 (as indicated in Figure 3), which is supported by multiple lines of evidence,

such as STS and BAC map data. As used herein, two proteins (or a region of the proteins) have significant homology when the amino acid sequences are typically at least about 70-80%, 80-90%, and more typically at least about 90-95% or more homologous. A significantly homologous amino acid sequence, according to the present invention, will be encoded by a nucleic acid sequence that will hybridize to a Ras-like protein polypeptide encoding nucleic acid molecule under stringent conditions as more fully described below.

Figure 3 provides information on SNPs that have been found in the gene encoding the Ras-like protein of the present invention. SNPs were identified at 25 different nucleotide positions. Some of these SNPs that are located outside the ORF and in introns may affect gene transcription.

Paralogs of a Ras-like protein polypeptide can readily be identified as having some degree of significant sequence homology/identity to at least a portion of the Ras-like protein polypeptide, as being encoded by a gene from humans, and as having similar activity or function. Two proteins will typically be considered paralogs when the amino acid sequences are typically at least about 40-50%, 50-60%, and more typically at least about 60-70% or more homologous through a given region or domain. Such paralogs will be encoded by a nucleic acid sequence that will hybridize to a Ras-like protein polypeptide encoding nucleic acid molecule under moderate to stringent conditions as more fully described below.

Orthologs of a Ras-like protein polypeptide can readily be identified as having some degree of significant sequence homology/identity to at least a portion of the Ras-like protein polypeptide as well as being encoded by a gene from another organism. Preferred orthologs will be isolated from mammals, preferably primates, for the development of human therapeutic targets and agents. Such orthologs will be encoded by a nucleic acid sequence that will hybridize to a Ras-like protein polypeptide encoding nucleic acid molecule under moderate to stringent conditions, as more fully described below, depending on the degree of relatedness of the two organisms yielding the proteins.

Non-naturally occurring variants of the Ras-like protein polypeptides of the present invention can readily be generated using recombinant techniques. Such variants include, but are not limited to deletions, additions and substitutions in the amino acid sequence of the

Ras-like protein polypeptide. For example, one class of substitutions is conserved amino acid substitutions. Such substitutions are those that substitute a given amino acid in a Ras-like protein polypeptide by another amino acid of like characteristics. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu, and Ile; interchange of the hydroxyl residues Ser and Thr, exchange of the acidic residues Asp and Glu, substitution between the amide residues Asn and Gln, exchange of the basic residues Lys and Arg, replacements among the aromatic residues Phe, Tyr, and the like. Guidance concerning which amino acid changes are likely to be phenotypically silent are found in Bowie *et al.*, *Science* 247:1306-1310 (1990).

Variant Ras-like protein polypeptides can be fully functional or can lack function in one or more activities. Fully functional variants typically contain only conservative variations or variations in non-critical residues or in non-critical regions. Functional variants can also contain substitution of similar amino acids that result in no change or an insignificant change in function. Alternatively, such substitutions may positively or negatively affect function to some degree.

Non-functional variants typically contain one or more non-conservative amino acid substitutions, deletions, insertions, inversions, or truncation or a substitution, insertion, inversion, or deletion in a critical residue or critical region.

Amino acids that are essential for function can be identified by methods known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham *et al.*, *Science* 244:1081-1085 (1989)). The latter procedure introduces single alanine mutations at every residue in the molecule. The resulting mutant molecules are then tested for biological activity such as receptor binding or *in vitro* proliferative activity. Sites that are critical for ligand-receptor binding can also be determined by structural analysis such as crystallography, nuclear magnetic resonance, or photoaffinity labeling (Smith *et al.*, *J. Mol. Biol.* 224:899-904 (1992); de Vos *et al.* *Science* 255:306-312 (1992)).

The present invention further provides fragments of the Ras-like protein polypeptides, in addition to proteins and peptides that comprise and consist of such fragments. Particularly those comprising the residues identified in Figure 2. The fragments to which the invention pertains, however, are not to be construed as encompassing fragments that have been disclosed publicly prior to the present invention.

As used herein, a fragment comprises at least 8, 10, 12, 14, 16 or more contiguous amino acid residues from a Ras-like protein polypeptide. Such fragments can be chosen based on the ability to retain one or more of the biological activities of the Ras-like protein polypeptide, or can be chosen for the ability to perform a function, e.g., act as an immunogen. Particularly important fragments are biologically active fragments, peptides that are, for example about 8 or more amino acids in length. Such fragments will typically comprise a domain or motif of the Ras-like protein polypeptide, e.g., active site. Further, possible fragments include, but are not limited to, domain or motif containing fragments, soluble peptide fragments, and fragments containing immunogenic structures. Predicted domains and functional sites are readily identifiable by computer programs well known and readily available to those of skill in the art (e.g., PROSITE, HMMer, eMOTIF, etc.). The results of one such analysis are provided in Figure 2.

Polypeptides often contain amino acids other than the 20 amino acids commonly referred to as the 20 naturally occurring amino acids. Further, many amino acids, including the terminal amino acids, may be modified by natural processes, such as processing and other post-translational modifications, or by chemical modification techniques well known in the art. Common modifications that occur naturally in Ras-like protein polypeptides are described in basic texts, detailed monographs, and the research literature, and they are well known to those of skill in the art (some of these features are identified in Figure 2).

Known modifications include, but are not limited to, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent crosslinks, formation of cystine, formation of pyroglutamate, formylation, gamma carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

Such modifications are well known to those of skill in the art and have been described in great detail in the scientific literature. Several particularly common

modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as *Proteins - Structure and Molecular Properties*, 2nd Ed., T.E. Creighton, W. H. Freeman and Company, New York (1993). Many detailed reviews are available on this subject, such as by Wold, F., *Posttranslational Covalent Modification of Proteins*, B.C. Johnson, Ed., Academic Press, New York 1-12 (1983); Seifter *et al.* (*Meth. Enzymol.* 182: 626-646 (1990)) and Rattan *et al.* (*Ann. N.Y. Acad. Sci.* 663:48-62 (1992)).

Accordingly, the Ras-like protein polypeptides of the present invention also encompass derivatives or analogs in which a substituted amino acid residue is not one encoded by the genetic code, in which a substituent group is included, in which the mature Ras-like protein polypeptide is fused with another compound, such as a compound to increase the half-life of the Ras-like protein polypeptide (for example, polyethylene glycol), or in which the additional amino acids are fused to the mature Ras-like protein polypeptide, such as a leader or secretory sequence or a sequence for purification of the mature Ras-like protein polypeptide, or a pro-protein sequence.

Protein/Peptide Uses

The proteins of the present invention can be used in assays to determine the biological activity of the protein, including in a panel of multiple proteins for high-throughput screening; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its ligand or receptor) in biological fluids; and as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in a disease state). Where the protein binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the protein can be used to identify the binding partner so as to develop a system to identify inhibitors of the binding interaction. Any or all of these research utilities are capable of being developed into reagent grade or kit format for commercialization as research products.

Methods for performing the uses listed above are well known to those skilled in the art. References disclosing such methods include "Molecular Cloning: A Laboratory Manual", 2d ed., Cold Spring Harbor Laboratory Press, Sambrook, J., E. F. Fritsch and T.

Maniatis eds., 1989, and "Methods in Enzymology: Guide to Molecular Cloning Techniques", Academic Press, Berger, S. L. and A. R. Kimmel eds., 1987.

The potential uses of the peptides of the present invention are based primarily on the source of the protein as well as the class/action of the protein. For example, Ras-like proteins isolated from humans and their human/mammalian orthologs serve as targets for identifying agents for use in mammalian therapeutic applications, e.g. a human drug, particularly in modulating a biological or pathological response in a cell or tissue that expresses the Ras-like protein. Experimental data as provided in Figure 1 indicates that Ras-like proteins of the present invention are expressed in humans in fetal brain, adrenal gland, brain neuroblastoma, skin melanotic melanoma, uterus leiomyosarcoma, and placenta, as indicated by virtual northern blot analysis. PCR-based tissue screening panels also indicate expression in the brain. A large percentage of pharmaceutical agents are being developed that modulate the activity of Ras-like proteins, particularly members of the Rab subfamily (see Background of the Invention). The structural and functional information provided in the Background and Figures provide specific and substantial uses for the molecules of the present invention, particularly in combination with the expression information provided in Figure 1. Experimental data as provided in Figure 1 indicates expression in the adult and fetal brain, adrenal gland, brain neuroblastoma, skin melanotic melanoma, uterus leiomyosarcoma, and placenta. Such uses can readily be determined using the information provided herein, that which is known in the art, and routine experimentation.

The proteins of the present invention (including variants and fragments that may have been disclosed prior to the present invention) are useful for biological assays related to Ras-like proteins that are related to members of the Rab subfamily. Such assays involve any of the known Ras-like protein functions or activities or properties useful for diagnosis and treatment of Ras-like protein-related conditions that are specific for the subfamily of Ras-like proteins that the one of the present invention belongs to, particularly in cells and tissues that express the Ras-like protein. Experimental data as provided in Figure 1 indicates that Ras-like proteins of the present invention are expressed in humans in fetal brain, adrenal gland, brain neuroblastoma, skin melanotic melanoma, uterus leiomyosarcoma, and

placenta, as indicated by virtual northern blot analysis. PCR-based tissue screening panels also indicate expression in the brain.

The proteins of the present invention are also useful in drug screening assays, in cell-based or cell-free systems. Cell-based systems can be native, i.e., cells that normally express the Ras-like protein, as a biopsy or expanded in cell culture. Experimental data as provided in Figure 1 indicates expression in the adult and fetal brain, adrenal gland, brain neuroblastoma, skin melanotic melanoma, uterus leiomyosarcoma, and placenta. In an alternate embodiment, cell-based assays involve recombinant host cells expressing the Ras-like protein.

The polypeptides can be used to identify compounds that modulate Ras-like protein activity. Both the Ras-like protein of the present invention and appropriate variants and fragments can be used in high-throughput screens to assay candidate compounds for the ability to bind to the Ras-like protein. These compounds can be further screened against a functional Ras-like protein to determine the effect of the compound on the Ras-like protein activity. Further, these compounds can be tested in animal or invertebrate systems to determine activity/effectiveness. Compounds can be identified that activate (agonist) or inactivate (antagonist) the Ras-like protein to a desired degree.

Therefore, in one embodiment, Rab or a fragment or derivative thereof may be administered to a subject to prevent or treat a disorder associated with an increase in apoptosis. Such disorders include, but are not limited to, AIDS and other infectious or genetic immunodeficiencies, neurodegenerative diseases such as Alzheimer's disease, Parkinson's disease, amyotrophic lateral sclerosis, retinitis pigmentosa, and cerebellar degeneration, myelodysplastic syndromes such as aplastic anemia, ischemic injuries such as myocardial infarction, stroke, and reperfusion injury, toxin-induced diseases such as alcohol-induced liver damage, cirrhosis, and lathyrism, wasting diseases such as cachexia, viral infections such as those caused by hepatitis B and C, and osteoporosis.

In another embodiment, a pharmaceutical composition comprising Rab may be administered to a subject to prevent or treat a disorder associated with increased apoptosis including, but not limited to, those listed above.

In still another embodiment, an agonist which is specific for Rab may be administered to prevent or treat a disorder associated with increased apoptosis including, but not limited to, those listed above.

In a further embodiment, a vector capable of expressing Rab, or a fragment or a derivative thereof, may be used to prevent or treat a disorder associated with increased apoptosis including, but not limited to, those listed above.

In cancer, where Rab promotes cell proliferation, it is desirable to decrease its activity. Therefore, in one embodiment, an antagonist of Rab may be administered to a subject to prevent or treat cancer including, but not limited to, adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, and teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. In one aspect, an antibody specific for Rab may be used directly as an antagonist, or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express Rab.

In another embodiment, a vector expressing the complement of the polynucleotide encoding Rab may be administered to a subject to prevent or treat a cancer including, but not limited to, the types of cancer listed above.

In inflammation, where Rab promotes cell proliferation, it is desirable to decrease its activity. Therefore, in one embodiment, an antagonist of Rab may be administered to a subject to prevent or treat an inflammation. Disorders associated with inflammation include, but are not limited to, Addison's disease, adult respiratory distress syndrome, allergies, anemia, asthma, atherosclerosis, bronchitis, cholecystitis, Crohn's disease, ulcerative colitis, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, atrophic gastritis, glomerulonephritis, gout, Graves' disease, hypereosinophilia, irritable bowel syndrome, lupus erythematosus, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, rheumatoid arthritis, scleroderma, Sjogren's syndrome, and autoimmune thyroiditis; complications of cancer, hemodialysis, extracorporeal circulation; viral, bacterial, fungal, parasitic, protozoal, and helminthic infections and trauma. In one aspect, an antibody specific for Rab may be used

directly as an antagonist, or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express Rab.

Further, the Ras-like protein polypeptides can be used to screen a compound for the ability to stimulate or inhibit interaction between the Ras-like protein and a molecule that normally interacts with the Ras-like protein, e.g. a ligand or a component of the signal pathway that the Ras-like protein normally interacts. Such assays typically include the steps of combining the Ras-like protein with a candidate compound under conditions that allow the Ras-like protein, or fragment, to interact with the target molecule, and to detect the formation of a complex between the protein and the target or to detect the biochemical consequence of the interaction with the Ras-like protein and the target, such as any of the associated effects of signal transduction.

Candidate compounds include, for example, 1) peptides such as soluble peptides, including Ig-tailed fusion peptides and members of random peptide libraries (see, e.g., Lam *et al.*, *Nature* 354:82-84 (1991); Houghten *et al.*, *Nature* 354:84-86 (1991)) and combinatorial chemistry-derived molecular libraries made of D- and/or L- configuration amino acids; 2) phosphopeptides (e.g., members of random and partially degenerate, directed phosphopeptide libraries, see, e.g., Songyang *et al.*, *Cell* 72:767-778 (1993)); 3) antibodies (e.g., polyclonal, monoclonal, humanized, anti-idiotypic, chimeric, and single chain antibodies as well as Fab, F(ab')₂, Fab expression library fragments, and epitope-binding fragments of antibodies); and 4) small organic and inorganic molecules (e.g., molecules obtained from combinatorial and natural product libraries). (Hodgson, *Bio/technology*, 1992, Sept 10(9);973-80).

One candidate compound is a soluble fragment of the Ras-like protein that competes for ligand binding. Other candidate compounds include mutant Ras-like proteins or appropriate fragments containing mutations that affect Ras-like protein function and thus compete for ligand. Accordingly, a fragment that competes for ligand, for example with a higher affinity, or a fragment that binds ligand but does not allow release, is within the scope of the invention.

The invention further includes other end point assays to identify compounds that modulate (stimulate or inhibit) Ras-like protein activity. The assays typically involve an assay of events in the Ras-like protein mediated signal transduction pathway that indicate

Ras-like protein activity. Thus, the phosphorylation of a protein/ligand target, the expression of genes that are up- or down-regulated in response to the Ras-like protein dependent signal cascade can be assayed. In one embodiment, the regulatory region of such genes can be operably linked to a marker that is easily detectable, such as luciferase. Alternatively, phosphorylation of the Ras-like protein, or a Ras-like protein target, could also be measured.

Any of the biological or biochemical functions mediated by the Ras-like protein can be used as an endpoint assay. These include all of the biochemical or biochemical/biological events described herein, in the references cited herein, incorporated by reference for these endpoint assay targets, and other functions known to those of ordinary skill in the art.

Binding and/or activating compounds can also be screened by using chimeric Ras-like proteins in which any of the protein's domains, or parts thereof, can be replaced by heterologous domains or subregions. Accordingly, a different set of signal transduction components is available as an end-point assay for activation. This allows for assays to be performed in other than the specific host cell from which the Ras-like protein is derived.

The Ras-like protein polypeptide of the present invention is also useful in competition binding assays in methods designed to discover compounds that interact with the Ras-like protein. Thus, a compound is exposed to a Ras-like protein polypeptide under conditions that allow the compound to bind or to otherwise interact with the polypeptide. Soluble Ras-like protein polypeptide is also added to the mixture. If the test compound interacts with the soluble Ras-like protein polypeptide, it decreases the amount of complex formed or activity from the Ras-like protein target. This type of assay is particularly useful in cases in which compounds are sought that interact with specific regions of the Ras-like protein. Thus, the soluble polypeptide that competes with the target Ras-like protein region is designed to contain peptide sequences corresponding to the region of interest.

To perform cell free drug screening assays, it is sometimes desirable to immobilize either the Ras-like protein, or fragment, or its target molecule to facilitate separation of complexes from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay.

Techniques for immobilizing proteins on matrices can be used in the drug screening assays. In one embodiment, a fusion protein can be provided which adds a domain that allows the protein to be bound to a matrix. For example, glutathione-S-transferase/15625 fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the cell lysates (e.g., ^{35}S -labeled) and the candidate compound, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads are washed to remove any unbound label, and the matrix immobilized and radiolabel determined directly, or in the supernatant after the complexes are dissociated. Alternatively, the complexes can be dissociated from the matrix, separated by SDS-PAGE, and the level of Ras-like protein-binding protein found in the bead fraction quantitated from the gel using standard electrophoretic techniques. For example, either the polypeptide or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin with techniques well known in the art. Alternatively, antibodies reactive with the protein but which do not interfere with binding of the protein to its target molecule can be derivatized to the wells of the plate, and the protein trapped in the wells by antibody conjugation. Preparations of a Ras-like protein-presenting wells and the amount of complex trapped in the well can be quantitated. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the Ras-like protein target molecule, or which are reactive with Ras-like protein and compete with the target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the target molecule.

Agents that modulate one of the Ras-like proteins of the present invention can be identified using one or more of the above assays, alone or in combination. It is generally preferable to use a cell-based or cell free system first and then confirm activity in an animal/insect model system. Such model systems are well known in the art and can readily be employed in this context.

Modulators of Ras-like protein activity identified according to these drug screening assays can be used to treat a subject with a disorder mediated by the Ras-like protein

associated pathway, by treating cells that express the Ras-like protein. Experimental data as provided in Figure 1 indicates expression in the adult and fetal brain, adrenal gland, brain neuroblastoma, skin melanotic melanoma, uterus leiomyosarcoma, and placenta. These methods of treatment include the steps of administering the modulators of protein activity in a pharmaceutical composition as described herein, to a subject in need of such treatment.

In yet another aspect of the invention, the Ras-like proteins can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos et al., *Cell* 72:223-232 (1993); Madura et al., *J. Biol. Chem.* 268:12046-12054 (1993); Bartel et al., *Biotechniques* 14:920-924 (1993); Iwabuchi et al., *Oncogene* 8:1693-1696 (1993); and Brent WO94/10300), to identify other proteins that bind to or interact with the Ras-like protein and are involved in Ras-like protein activity. Such Ras-like protein-binding proteins are also likely to be involved in the propagation of signals by the Ras-like proteins or Ras-like protein targets as, for example, downstream elements of a Ras-like protein-mediated signaling pathway, e.g., a pain signaling pathway. Alternatively, such Ras-like protein-binding proteins are likely to be Ras-like protein inhibitors.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for a Ras-like protein is fused to a gene encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming a Ras-like protein-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with the Ras-like protein.

This invention further pertains to novel agents identified by the above-described screening assays. Accordingly, it is within the scope of this invention to further use an agent identified as described herein in an appropriate animal model. For example, an agent identified as described herein (e.g., a Ras-like protein modulating agent, an antisense Ras-like protein nucleic acid molecule, a Ras-like protein-specific antibody, or a Ras-like protein-binding partner) can be used in an animal or insect model to determine the efficacy, toxicity, or side effects of treatment with such an agent. Alternatively, an agent identified as described herein can be used in an animal or insect model to determine the mechanism of action of such an agent. Furthermore, this invention pertains to uses of novel agents identified by the above-described screening assays for treatments as described herein.

The Ras-like proteins of the present invention are also useful to provide a target for diagnosing a disease or predisposition to a disease mediated by the peptide. Accordingly, the invention provides methods for detecting the presence, or levels of, the protein (or encoding mRNA) in a cell, tissue, or organism. Experimental data as provided in Figure 1 indicates expression in the adult and fetal brain, adrenal gland, brain neuroblastoma, skin melanotic melanoma, uterus leiomyosarcoma, and placenta. The method involves contacting a biological sample with a compound capable of interacting with the receptor protein such that the interaction can be detected. Such an assay can be provided in a single detection format or a multi-detection format such as an antibody chip array.

One agent for detecting a protein in a sample is an antibody capable of selectively binding to protein. A biological sample includes tissues, cells and biological fluids isolated from a subject, as well as tissues, cells, and fluids present within a subject.

The peptides also are useful to provide a target for diagnosing a disease or predisposition to a disease mediated by the peptide. Accordingly, the invention provides methods for detecting the presence, or levels of, the protein in a cell, tissue, or organism. The method involves contacting a biological sample with a compound capable of interacting with the receptor protein such that the interaction can be detected.

The peptides of the present invention also provide targets for diagnosing active disease, or predisposition to a disease, in a patient having a variant peptide. Thus, the peptide can be isolated from a biological sample and assayed for the presence of a genetic

mutation that results in translation of an aberrant peptide. This includes amino acid substitution, deletion, insertion, rearrangement, (as the result of aberrant splicing events), and inappropriate post-translational modification. Analytic methods include altered electrophoretic mobility, altered tryptic peptide digest, altered receptor activity in cell-based or cell-free assay, alteration in ligand or antibody-binding pattern, altered isoelectric point, direct amino acid sequencing, and any other of the known assay techniques useful for detecting mutations in a protein. Such an assay can be provided in a single detection format or a multi-detection format such as an antibody chip array.

In vitro techniques for detection of peptide include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations, and immunofluorescence using a detection reagents, such as an antibody or protein binding agent.. Alternatively, the peptide can be detected *in vivo* in a subject by introducing into the subject a labeled anti-peptide antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques. Particularly useful are methods that detect the allelic variant of a peptide expressed in a subject and methods which detect fragments of a peptide in a sample.

The peptides are also useful in pharmacogenomic analysis. Pharmacogenomics deal with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See, e.g., Eichelbaum, M. (*Clin. Exp. Pharmacol. Physiol.* 23(10-11) :983-985 (1996)), and Linder, M.W. (*Clin. Chem.* 43(2):254-266 (1997)). The clinical outcomes of these variations result in severe toxicity of therapeutic drugs in certain individuals or therapeutic failure of drugs in certain individuals as a result of individual variation in metabolism. Thus, the genotype of the individual can determine the way a therapeutic compound acts on the body or the way the body metabolizes the compound. Further, the activity of drug metabolizing enzymes effects both the intensity and duration of drug action. Thus, the pharmacogenomics of the individual permit the selection of effective compounds and effective dosages of such compounds for prophylactic or therapeutic treatment based on the individual's genotype. The discovery of genetic polymorphisms in some drug metabolizing enzymes has explained why some patients do not obtain the expected drug effects, show an exaggerated drug effect, or experience serious toxicity from standard drug dosages. Polymorphisms can be expressed

in the phenotype of the extensive metabolizer and the phenotype of the poor metabolizer. Accordingly, genetic polymorphism may lead to allelic protein variants of the receptor protein in which one or more of the receptor functions in one population is different from those in another population. The peptides thus allow a target to ascertain a genetic predisposition that can affect treatment modality. Thus, in a ligand-based treatment, polymorphism may give rise to amino terminal extracellular domains and/or other ligand-binding regions that are more or less active in ligand binding, and receptor activation. Accordingly, ligand dosage would necessarily be modified to maximize the therapeutic effect within a given population containing a polymorphism. As an alternative to genotyping, specific polymorphic peptides could be identified.

The peptides are also useful for treating a disorder characterized by an absence of, inappropriate, or unwanted expression of the protein. Experimental data as provided in Figure 1 indicates expression in the adult and fetal brain, adrenal gland, brain neuroblastoma, skin melanotic melanoma, uterus leiomyosarcoma, and placenta. Accordingly, methods for treatment include the use of the Ras-like protein or fragments.

Antibodies

The invention also provides antibodies that selectively bind to one of the peptides of the present invention, a protein comprising such a peptide, as well as variants and fragments thereof. As used herein, an antibody selectively binds a target peptide when it binds the target peptide and does not significantly bind to unrelated proteins. An antibody is still considered to selectively bind a peptide even if it also binds to other proteins that are not substantially homologous with the target peptide so long as such proteins share homology with a fragment or domain of the peptide target of the antibody. In this case, it would be understood that antibody binding to the peptide is still selective despite some degree of cross-reactivity.

As used herein, an antibody is defined in terms consistent with that recognized within the art: they are multi-subunit proteins produced by a mammalian organism in response to an antigen challenge. The antibodies of the present invention include polyclonal antibodies and monoclonal antibodies, as well as fragments of such antibodies, including, but not limited to, Fab or F(ab')₂, and Fv fragments.

Many methods are known for generating and/or identifying antibodies to a given target peptide. Several such methods are described by Harlow, Antibodies, Cold Spring Harbor Press, (1989).

In general, to generate antibodies, an isolated peptide is used as an immunogen and is administered to a mammalian organism, such as a rat, rabbit or mouse. The full-length protein, an antigenic peptide fragment or a fusion protein can be used. Particularly important fragments are those covering functional domains, such as the domains identified in Figure 2, and domain of sequence homology or divergence amongst the family, such as those that can readily be identified using protein alignment methods and as presented in the Figures.

Antibodies are preferably prepared from regions or discrete fragments of the Ras-like proteins. Antibodies can be prepared from any region of the peptide as described herein. However, preferred regions will include those involved in function/activity and/or receptor/binding partner interaction. Figure 2 can be used to identify particularly important regions while sequence alignment can be used to identify conserved and unique sequence fragments.

An antigenic fragment will typically comprise at least 8 contiguous amino acid residues. The antigenic peptide can comprise, however, at least 10, 12, 14, 16 or more amino acid residues. Such fragments can be selected on a physical property, such as fragments correspond to regions that are located on the surface of the protein, e.g., hydrophilic regions or can be selected based on sequence uniqueness (see Figure 2)..

Detection of an antibody of the present invention can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes

luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I , ^{131}I , ^{35}S , or ^3H .

Antibody Uses

The antibodies can be used to isolate one of the proteins of the present invention by standard techniques, such as affinity chromatography or immunoprecipitation. The antibodies can facilitate the purification of the natural protein from cells and recombinantly produced protein expressed in host cells. In addition, such antibodies are useful to detect the presence of one of the proteins of the present invention in cells or tissues to determine the pattern of expression of the protein among various tissues in an organism and over the course of normal development. Experimental data as provided in Figure 1 indicates that Ras-like proteins of the present invention are expressed in humans in fetal brain, adrenal gland, brain neuroblastoma, skin melanotic melanoma, uterus leiomyosarcoma, and placenta, as indicated by virtual northern blot analysis. PCR-based tissue screening panels also indicate expression in the brain. Further, such antibodies can be used to detect protein *in situ*, *in vitro*, or in a cell lysate or supernatant in order to evaluate the abundance and pattern of expression. Also, such antibodies can be used to assess abnormal tissue distribution or abnormal expression during development. Antibody detection of circulating fragments of the full-length protein can be used to identify turnover.

Further, the antibodies can be used to assess expression in disease states such as in active stages of the disease or in an individual with a predisposition toward disease related to the protein's function. When a disorder is caused by an inappropriate tissue distribution, developmental expression, level of expression of the protein, or expressed/processed form, the antibody can be prepared against the normal protein. Experimental data as provided in Figure 1 indicates expression in the adult and fetal brain, adrenal gland, brain neuroblastoma, skin melanotic melanoma, uterus leiomyosarcoma, and placenta. If a disorder is characterized by a specific mutation in the protein, antibodies specific for this mutant protein can be used to assay for the presence of the specific mutant protein.

The antibodies can also be used to assess normal and aberrant subcellular localization of cells in the various tissues in an organism. Experimental data as provided in Figure 1 indicates expression in the adult and fetal brain, adrenal gland, brain neuroblastoma, skin melanotic melanoma, uterus leiomyosarcoma, and placenta. The

diagnostic uses can be applied, not only in genetic testing, but also in monitoring a treatment modality. Accordingly, where treatment is ultimately aimed at correcting expression level or the presence of aberrant sequence and aberrant tissue distribution or developmental expression, antibodies directed against the or relevant fragments can be used to monitor therapeutic efficacy.

Additionally, antibodies are useful in pharmacogenomic analysis. Thus, antibodies prepared against polymorphic proteins can be used to identify individuals that require modified treatment modalities. The antibodies are also useful as diagnostic tools as an immunological marker for aberrant protein analyzed by electrophoretic mobility, isoelectric point, tryptic peptide digest, and other physical assays known to those in the art.

The antibodies are also useful for tissue typing. Experimental data as provided in Figure 1 indicates expression in the adult and fetal brain, adrenal gland, brain neuroblastoma, skin melanotic melanoma, uterus leiomyosarcoma, and placenta. Thus, where a specific protein has been correlated with expression in a specific tissue, antibodies that are specific for this protein can be used to identify a tissue type.

The antibodies are also useful for inhibiting protein function, for example, blocking the binding of the Ras-like protein to a binding partner such as a substrate. These uses can also be applied in a therapeutic context in which treatment involves inhibiting the protein's function. An antibody can be used, for example, to block binding, thus modulating (agonizing or antagonizing) the peptides activity. Antibodies can be prepared against specific fragments containing sites required for function or against intact protein that is associated with a cell or cell membrane. See Figure 2 for structural information relating to the proteins of the present invention.

The invention also encompasses kits for using antibodies to detect the presence of a protein in a biological sample. The kit can comprise antibodies such as a labeled or labelable antibody and a compound or agent for detecting protein in a biological sample; means for determining the amount of protein in the sample; means for comparing the amount of protein in the sample with a standard; and instructions for use.

Nucleic Acid Molecules

The present invention further provides isolated nucleic acid molecules that encode a Ras-like protein polypeptide of the present invention. Such nucleic acid molecules will

consist of, consist essentially of, or comprise a nucleotide sequence that encodes one of the Ras-like protein polypeptides of the present invention, an allelic variant thereof, or an ortholog or paralog thereof.

As used herein, an "isolated" nucleic acid molecule is one that is separated from other nucleic acid present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. However, there can be some flanking nucleotide sequences, for example up to about 5KB, particularly contiguous peptide encoding sequences and peptide encoding sequences within the same gene but separated by introns in the genomic sequence. The important point is that the nucleic acid is isolated from remote and unimportant flanking sequences such that it can be subjected to the specific manipulations described herein such as recombinant expression, preparation of probes and primers, and other uses specific to the nucleic acid sequences.

Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized. However, the nucleic acid molecule can be fused to other coding or regulatory sequences and still be considered isolated.

For example, recombinant DNA molecules contained in a vector are considered isolated. Further examples of isolated DNA molecules include recombinant DNA molecules maintained in heterologous host cells or purified (partially or substantially) DNA molecules in solution. Isolated RNA molecules include *in vivo* or *in vitro* RNA transcripts of the isolated DNA molecules of the present invention. Isolated nucleic acid molecules according to the present invention further include such molecules produced synthetically.

Accordingly, the present invention provides nucleic acid molecules that consist of the nucleotide sequence shown in Figure 1 or 3 (SEQ ID NO:1, transcript sequence and SEQ ID NO:3, genomic sequence), or any nucleic acid molecule that encodes the protein provided in Figure 2, SEQ ID NO:2. A nucleic acid molecule consists of a nucleotide sequence when the nucleotide sequence is the complete nucleotide sequence of the nucleic acid molecule. The present invention further provides nucleic acid molecules that consist

essentially of the nucleotide sequence shown in Figure 1 or 3 (SEQ ID NO:1, transcript sequence and SEQ ID NO:3, genomic sequence), or any nucleic acid molecule that encodes the protein provided in Figure 2, SEQ ID NO:2. A nucleic acid molecule consists essentially of a nucleotide sequence when such a nucleotide sequence is present with only a few additional nucleic acid residues in the final nucleic acid molecule.

The present invention further provides nucleic acid molecules that comprise the nucleotide sequences shown in Figure 1 or 3 (SEQ ID NO:1, transcript sequence and SEQ ID NO:3, genomic sequence), or any nucleic acid molecule that encodes the protein provided in Figure 2, SEQ ID NO:2. A nucleic acid molecule comprises a nucleotide sequence when the nucleotide sequence is at least part of the final nucleotide sequence of the nucleic acid molecule. In such a fashion, the nucleic acid molecule can be only the nucleotide sequence or have additional nucleic acid residues, such as nucleic acid residues that are naturally associated with it or heterologous nucleotide sequences. Such a nucleic acid molecule can have a few additional nucleotides or can comprises several hundred or more additional nucleotides. A brief description of how various types of these nucleic acid molecules can be readily made/isolated is provided below.

In Figures 1 and 3, both coding and non-coding sequences are provided. Because of the source of the present invention, humans genomic sequence (Figure 3) and cDNA/transcript sequences (Figure 1), the nucleic acid molecules in the Figures will contain genomic intronic sequences, 5' and 3' non-coding sequences, gene regulatory regions and non-coding intergenic sequences. In general such sequence features are either noted in Figures 1 and 3 or can readily be identified using computational tools known in the art. As discussed below, some of the non-coding regions, particularly gene regulatory elements such as promoters, are useful for a variety of purposes, e.g. control of heterologous gene expression, target for identifying gene activity modulating compounds, and are particularly claimed as fragments of the genomic sequence provided herein.

Full-length genes may be cloned from known sequence using any one of a number of methods known in the art. For example, a method which employs XL-PCR (Perkin-Elmer, Foster City, Calif.) to amplify long pieces of DNA may be used. Other methods for obtaining full-length sequences are well known in the art.

The isolated nucleic acid molecules can encode the mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids interior to the mature peptide (when the mature form has more than one peptide chain, for instance). Such sequences may play a role in processing of a protein from precursor to a mature form, facilitate protein trafficking, prolong or shorten protein half-life, or facilitate manipulation of a protein for assay or production, among other things. As generally is the case *in situ*, the additional amino acids may be processed away from the mature protein by cellular enzymes.

As mentioned above, the isolated nucleic acid molecules include, but are not limited to, the sequence encoding the Ras-like protein polypeptide alone, the sequence encoding the mature peptide and additional coding sequences, such as a leader or secretory sequence (e.g., a pre-pro or pro-protein sequence), the sequence encoding the mature peptide, with or without the additional coding sequences, plus additional non-coding sequences, for example introns and non-coding 5' and 3' sequences such as transcribed but non-translated sequences that play a role in transcription, mRNA processing (including splicing and polyadenylation signals), ribosome binding, and stability of mRNA. In addition, the nucleic acid molecule may be fused to a marker sequence encoding, for example, a peptide that facilitates purification.

Isolated nucleic acid molecules can be in the form of RNA, such as mRNA, or in the form of DNA, including cDNA and genomic DNA obtained by cloning or produced by chemical synthetic techniques or by a combination thereof. The nucleic acid, especially DNA, can be double-stranded or single-stranded. Single-stranded nucleic acid can be the coding strand (sense strand) or the non-coding strand (anti-sense strand).

The invention further provides nucleic acid molecules that encode fragments of the peptides of the present invention and that encode obvious variants of the Ras-like proteins of the present invention that are described above. Such nucleic acid molecules may be naturally occurring, such as allelic variants (same locus), paralogs (different locus), and orthologs (different organism), or may be constructed by recombinant DNA methods or by chemical synthesis. Such non-naturally occurring variants may be made by mutagenesis techniques, including those applied to nucleic acid molecules, cells, or whole organisms. Accordingly, as discussed above, the variants can contain nucleotide substitutions, deletions, inversions, and/or insertions. Variation can occur in either or both the coding and non-

coding regions. The variations can produce both conservative and non-conservative amino acid substitutions.

The present invention further provides non-coding fragments of the nucleic acid molecules provided in the Figures 1 and 3. Preferred non-coding fragments include, but are not limited to, promoter sequences, enhancer sequences, gene modulating sequences, and gene termination sequences. Such fragments are useful in controlling heterologous gene expression and in developing screens to identify gene-modulating agents.

A fragment comprises a contiguous nucleotide sequence greater than 12 or more nucleotides. Further, a fragment could be at least 30, 40, 50, 100 250, or 500 nucleotides in length. The length of the fragment will be based on its intended use. For example, the fragment can encode epitope-bearing regions of the peptide, or can be useful as DNA probes and primers. Such fragments can be isolated using the known nucleotide sequence to synthesize an oligonucleotide probe. A labeled probe can then be used to screen a cDNA library, genomic DNA library, or mRNA to isolate nucleic acid corresponding to the coding region. Further, primers can be used in PCR reactions to clone specific regions of gene.

A probe/primer typically comprises substantially a purified oligonucleotide or oligonucleotide pair. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 20, 25, 40, 50, or more consecutive nucleotides.

Orthologs, homologs, and allelic variants can be identified using methods well known in the art. As described in the Peptide Section, these variants comprise a nucleotide sequence encoding a peptide that is typically 60-70%, 70-80%, 80-90%, and more typically at least about 90-95% or more homologous to the nucleotide sequence shown in the Figure sheets or a fragment of this sequence. Such nucleic acid molecules can readily be identified as being able to hybridize under moderate to stringent conditions, to the nucleotide sequence shown in the Figure sheets or a fragment of the sequence. The gene encoding the novel Ras-like protein of the present invention is located on a genome component that has been mapped to human chromosome 2 (as indicated in Figure 3), which is supported by multiple lines of evidence, such as STS and BAC map data.

Figure 3 provides information on SNPs that have been found in the gene encoding the Ras-like protein of the present invention. SNPs were identified at 25 different nucleotide

positions. Some of these SNPs that are located outside the ORF and in introns may affect gene transcription.

As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences encoding a peptide at least 60-70% homologous to each other typically remain hybridized to each other. The conditions can be such that sequences at least about 60%, at least about 70%, or at least about 80% or more homologous to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. One example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65°C. Examples of moderate to low stringency hybridization conditions are well known in the art.

Nucleic Acid Molecule Uses

The nucleic acid molecules of the present invention are useful for probes, primers, chemical intermediates, and in biological assays. The nucleic acid molecules are useful as a hybridization probe for messenger RNA, transcript/cDNA and genomic DNA to isolate full-length cDNA and genomic clones encoding the peptide described in Figure 2 and to isolate cDNA and genomic clones that correspond to variants (alleles, orthologs, etc.) producing the same or related peptides shown in Figure 2. As illustrated in Figure 3, SNPs were identified at 25 different nucleotide positions.

The probe can correspond to any sequence along the entire length of the nucleic acid molecules provided in the Figures. Accordingly, it could be derived from 5' noncoding regions, the coding region, and 3' noncoding regions. However, as discussed, fragments are not to be construed as those, which may encompass fragments disclosed prior to the present invention.

The nucleic acid molecules are also useful as primers for PCR to amplify any given region of a nucleic acid molecule and are useful to synthesize antisense molecules of desired length and sequence.

The nucleic acid molecules are also useful for constructing recombinant vectors. Such vectors include expression vectors that express a portion of, or all of, the peptide sequences. Vectors also include insertion vectors, used to integrate into another nucleic acid molecule sequence, such as into the cellular genome, to alter *in situ* expression of a gene and/or gene product. For example, an endogenous coding sequence can be replaced via homologous recombination with all or part of the coding region containing one or more specifically introduced mutations.

The nucleic acid molecules are also useful for expressing antigenic portions of the proteins.

The nucleic acid molecules are also useful as probes for determining the chromosomal positions of the nucleic acid molecules by means of *in situ* hybridization methods. The gene encoding the novel Ras-like protein of the present invention is located on a genome component that has been mapped to human chromosome 2 (as indicated in Figure 3), which is supported by multiple lines of evidence, such as STS and BAC map data.

The nucleic acid molecules are also useful in making vectors containing the gene regulatory regions of the nucleic acid molecules of the present invention.

The nucleic acid molecules are also useful for designing ribozymes corresponding to all, or a part, of the mRNA produced from the nucleic acid molecules described herein.

The nucleic acid molecules are also useful for constructing host cells expressing a part, or all, of the nucleic acid molecules and peptides. Moreover, the nucleic acid molecules are useful for constructing transgenic animals wherein a homolog of the nucleic acid molecule has been “knocked-out” of the animal’s genome.

The nucleic acid molecules are also useful for constructing transgenic animals expressing all, or a part, of the nucleic acid molecules and peptides.

The nucleic acid molecules are also useful for making vectors that express part, or all, of the peptides.

The nucleic acid molecules are also useful as hybridization probes for determining the presence, level, form, and distribution of nucleic acid expression. Experimental data as provided in Figure 1 indicates that Ras-like proteins of the present invention are expressed in humans in fetal brain, adrenal gland, brain neuroblastoma, skin melanotic melanoma, uterus leiomyosarcoma, and placenta, as indicated by virtual northern blot analysis. PCR-

based tissue screening panels also indicate expression in the brain. Accordingly, the probes can be used to detect the presence of, or to determine levels of, a specific nucleic acid molecule in cells, tissues, and in organisms. The nucleic acid whose level is determined can be DNA or RNA. Accordingly, probes corresponding to the peptides described herein can be used to assess expression and/or gene copy number in a given cell, tissue, or organism. These uses are relevant for diagnosis of disorders involving an increase or decrease in Ras-like protein expression relative to normal results.

In vitro techniques for detection of mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detecting DNA include Southern hybridizations and *in situ* hybridization.

Probes can be used as a part of a diagnostic test kit for identifying cells or tissues that express a Ras-like protein, such as by measuring a level of a receptor-encoding nucleic acid in a sample of cells from a subject e.g., mRNA or genomic DNA, or determining if a receptor gene has been mutated. Experimental data as provided in Figure 1 indicates that Ras-like proteins of the present invention are expressed in humans in fetal brain, adrenal gland, brain neuroblastoma, skin melanotic melanoma, uterus leiomyosarcoma, and placenta, as indicated by virtual northern blot analysis. PCR-based tissue screening panels also indicate expression in the brain.

Nucleic acid expression assays are useful for drug screening to identify compounds that modulate Ras-like protein nucleic acid expression.

The invention thus provides a method for identifying a compound that can be used to treat a disorder associated with nucleic acid expression of the Ras-like protein gene, particularly biological and pathological processes that are mediated by the Ras-like protein in cells and tissues that express it. Experimental data as provided in Figure 1 indicates expression in the adult and fetal brain, adrenal gland, brain neuroblastoma, skin melanotic melanoma, uterus leiomyosarcoma, and placenta. The method typically includes assaying the ability of the compound to modulate the expression of the Ras-like protein nucleic acid and thus identifying a compound that can be used to treat a disorder characterized by undesired Ras-like protein nucleic acid expression. The assays can be performed in cell-based and cell-free systems. Cell-based assays include cells naturally expressing the Ras-

like protein nucleic acid or recombinant cells genetically engineered to express specific nucleic acid sequences.

The assay for Ras-like protein nucleic acid expression can involve direct assay of nucleic acid levels, such as mRNA levels, or on collateral compounds involved in the signal pathway. Further, the expression of genes that are up- or down-regulated in response to the Ras-like protein signal pathway can also be assayed. In this embodiment the regulatory regions of these genes can be operably linked to a reporter gene such as luciferase.

Thus, modulators of Ras-like protein gene expression can be identified in a method wherein a cell is contacted with a candidate compound and the expression of mRNA determined. The level of expression of Ras-like protein mRNA in the presence of the candidate compound is compared to the level of expression of Ras-like protein mRNA in the absence of the candidate compound. The candidate compound can then be identified as a modulator of nucleic acid expression based on this comparison and be used, for example to treat a disorder characterized by aberrant nucleic acid expression. When expression of mRNA is statistically significantly greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of nucleic acid expression. When nucleic acid expression is statistically significantly less in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of nucleic acid expression.

The invention further provides methods of treatment, with the nucleic acid as a target, using a compound identified through drug screening as a gene modulator to modulate Ras-like protein nucleic acid expression in cells and tissues that express the Ras-like protein. Experimental data as provided in Figure 1 indicates that Ras-like proteins of the present invention are expressed in humans in fetal brain, adrenal gland, brain neuroblastoma, skin melanotic melanoma, uterus leiomyosarcoma, and placenta, as indicated by virtual northern blot analysis. PCR-based tissue screening panels also indicate expression in the brain. Modulation includes both up-regulation (i.e. activation or agonization) or down-regulation (suppression or antagonization) of nucleic acid expression.

Alternatively, a modulator for Ras-like protein nucleic acid expression can be a small molecule or drug identified using the screening assays described herein as long as the drug or small molecule inhibits the Ras-like protein nucleic acid expression in the cells and

tissues that express the protein. Experimental data as provided in Figure 1 indicates expression in the adult and fetal brain, adrenal gland, brain neuroblastoma, skin melanotic melanoma, uterus leiomyosarcoma, and placenta.

The nucleic acid molecules are also useful for monitoring the effectiveness of modulating compounds on the expression or activity of the Ras-like protein gene in clinical trials or in a treatment regimen. Thus, the gene expression pattern can serve as a barometer for the continuing effectiveness of treatment with the compound, particularly with compounds to which a patient can develop resistance. The gene expression pattern can also serve as a marker indicative of a physiological response of the affected cells to the compound. Accordingly, such monitoring would allow either increased administration of the compound or the administration of alternative compounds to which the patient has not become resistant. Similarly, if the level of nucleic acid expression falls below a desirable level, administration of the compound could be commensurately decreased.

The nucleic acid molecules are also useful in diagnostic assays for qualitative changes in Ras-like protein nucleic acid, and particularly in qualitative changes that lead to pathology. The nucleic acid molecules can be used to detect mutations in Ras-like protein genes and gene expression products such as mRNA. The nucleic acid molecules can be used as hybridization probes to detect naturally occurring genetic mutations in the Ras-like protein gene and thereby to determine whether a subject with the mutation is at risk for a disorder caused by the mutation. Mutations include deletion, addition, or substitution of one or more nucleotides in the gene, chromosomal rearrangement, such as inversion or transposition, modification of genomic DNA, such as aberrant methylation patterns, or changes in gene copy number, such as amplification. Detection of a mutated form of the Ras-like protein gene associated with a dysfunction provides a diagnostic tool for an active disease or susceptibility to disease when the disease results from overexpression, underexpression, or altered expression of a Ras-like protein.

Individuals carrying mutations in the Ras-like protein gene can be detected at the nucleic acid level by a variety of techniques. Figure 3 provides information on SNPs that have been found in the gene encoding the Ras-like protein of the present invention. SNPs were identified at 25 different nucleotide positions. Some of these SNPs that are located outside the ORF and in introns may affect gene transcription. The gene encoding the novel

Ras-like protein of the present invention is located on a genome component that has been mapped to human chromosome 2 (as indicated in Figure 3), which is supported by multiple lines of evidence, such as STS and BAC map data. Genomic DNA can be analyzed directly or can be amplified by using PCR prior to analysis. RNA or cDNA can be used in the same way. In some uses, detection of the mutation involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, e.g. U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran *et al.*, *Science* 241:1077-1080 (1988); and Nakazawa *et al.*, *PNAS* 91:360-364 (1994)), the latter of which can be particularly useful for detecting point mutations in the gene (see Abravaya *et al.*, *Nucleic Acids Res.* 23:675-682 (1995)). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a gene under conditions such that hybridization and amplification of the gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. Deletions and insertions can be detected by a change in size of the amplified product compared to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to normal RNA or antisense DNA sequences.

Alternatively, mutations in a Ras-like protein gene can be directly identified, for example, by alterations in restriction enzyme digestion patterns determined by gel electrophoresis.

Further, sequence-specific ribozymes (U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site. Perfectly matched sequences can be distinguished from mismatched sequences by nuclease cleavage digestion assays or by differences in melting temperature.

Sequence changes at specific locations can also be assessed by nuclease protection assays such as RNase and S1 protection or the chemical cleavage method. Furthermore, sequence differences between a mutant Ras-like protein gene and a wild-type gene can be determined by direct DNA sequencing. A variety of automated sequencing procedures can be utilized when performing the diagnostic assays (Naeve, C.W., *Biotechniques* 19:448

(1995)), including sequencing by mass spectrometry (see, e.g., PCT International Publication No. WO 94/16101; Cohen *et al.*, *Adv. Chromatogr.* 36:127-162 (1996); and Griffin *et al.*, *Appl. Biochem. Biotechnol.* 38:147-159 (1993)).

Other methods for detecting mutations in the gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA duplexes (Myers *et al.*, *Science* 230:1242 (1985)); Cotton *et al.*, *PNAS* 85:4397 (1988); Saleeba *et al.*, *Meth. Enzymol.* 217:286-295 (1992)), electrophoretic mobility of mutant and wild type nucleic acid is compared (Orita *et al.*, *PNAS* 86:2766 (1989); Cotton *et al.*, *Mutat. Res.* 285:125-144 (1993); and Hayashi *et al.*, *Genet. Anal. Tech. Appl.* 9:73-79 (1992)), and movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (Myers *et al.*, *Nature* 313:495 (1985)). Examples of other techniques for detecting point mutations include, selective oligonucleotide hybridization, selective amplification, and selective primer extension.

The nucleic acid molecules are also useful for testing an individual for a genotype that while not necessarily causing the disease, nevertheless affects the treatment modality. Thus, the nucleic acid molecules can be used to study the relationship between an individual's genotype and the individual's response to a compound used for treatment (pharmacogenomic relationship). Accordingly, the nucleic acid molecules described herein can be used to assess the mutation content of the Ras-like protein gene in an individual in order to select an appropriate compound or dosage regimen for treatment. Figure 3 provides information on SNPs that have been found in the gene encoding the Ras-like protein of the present invention. SNPs were identified at 25 different nucleotide positions. Some of these SNPs that are located outside the ORF and in introns may affect gene transcription.

Thus nucleic acid molecules displaying genetic variations that affect treatment provide a diagnostic target that can be used to tailor treatment in an individual. Accordingly, the production of recombinant cells and animals containing these polymorphisms allow effective clinical design of treatment compounds and dosage regimens.

The nucleic acid molecules are thus useful as antisense constructs to control Ras-like protein gene expression in cells, tissues, and organisms. A DNA antisense nucleic acid

molecule is designed to be complementary to a region of the gene involved in transcription, preventing transcription and hence production of Ras-like protein. An antisense RNA or DNA nucleic acid molecule would hybridize to the mRNA and thus block translation of mRNA into Ras-like protein.

Alternatively, a class of antisense molecules can be used to inactivate mRNA in order to decrease expression of Ras-like protein nucleic acid. Accordingly, these molecules can treat a disorder characterized by abnormal or undesired Ras-like protein nucleic acid expression. This technique involves cleavage by means of ribozymes containing nucleotide sequences complementary to one or more regions in the mRNA that attenuate the ability of the mRNA to be translated. Possible regions include coding regions and particularly coding regions corresponding to the catalytic and other functional activities of the Ras-like protein, such as ligand binding.

The nucleic acid molecules also provide vectors for gene therapy in patients containing cells that are aberrant in Ras-like protein gene expression. Thus, recombinant cells, which include the patient's cells that have been engineered *ex vivo* and returned to the patient, are introduced into an individual where the cells produce the desired Ras-like protein to treat the individual.

The invention also encompasses kits for detecting the presence of a Ras-like protein nucleic acid in a biological sample. Experimental data as provided in Figure 1 indicates that Ras-like proteins of the present invention are expressed in humans in fetal brain, adrenal gland, brain neuroblastoma, skin melanotic melanoma, uterus leiomyosarcoma, and placenta, as indicated by virtual northern blot analysis. PCR-based tissue screening panels also indicate expression in the brain. For example, the kit can comprise reagents such as a labeled or labelable nucleic acid or agent capable of detecting Ras-like protein nucleic acid in a biological sample; means for determining the amount of Ras-like protein nucleic acid in the sample; and means for comparing the amount of Ras-like protein nucleic acid in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect Ras-like protein mRNA or DNA.

Nucleic Acid Arrays

The present invention further provides arrays or microarrays of nucleic acid molecules that are based on the sequence information provided in Figures 1 and 3 (SEQ ID NOS:1 and 3).

As used herein "Arrays" or "Microarrays" refers to an array of distinct polynucleotides or oligonucleotides synthesized on a substrate, such as paper, nylon or other type of membrane, filter, chip, glass slide, or any other suitable solid support. In one embodiment, the microarray is prepared and used according to the methods described in US Patent 5,837,832, Chee et al., PCT application W095/11995 (Chee et al.), Lockhart, D. J. et al. (1996; Nat. Biotech. 14: 1675-1680) and Schena, M. et al. (1996; Proc. Natl. Acad. Sci. 93: 10614-10619), all of which are incorporated herein in their entirety by reference. In other embodiments, such arrays are produced by the methods described by Brown et. al., US Patent No. 5,807,522.

The microarray is preferably composed of a large number of unique, single-stranded nucleic acid sequences, usually either synthetic antisense oligonucleotides or fragments of cDNAs, fixed to a solid support. The oligonucleotides are preferably about 6-60 nucleotides in length, more preferably 15-30 nucleotides in length, and most preferably about 20-25 nucleotides in length. For a certain type of microarray, it may be preferable to use oligonucleotides that are only 7-20 nucleotides in length. The microarray may contain oligonucleotides that cover the known 5', or 3', sequence, sequential oligonucleotides that cover the full-length sequence; or unique oligonucleotides selected from particular areas along the length of the sequence. Polynucleotides used in the microarray may be oligonucleotides that are specific to a gene or genes of interest.

In order to produce oligonucleotides to a known sequence for a microarray, the gene(s) of interest (or an ORF identified from the contigs of the present invention) is typically examined using a computer algorithm that starts at the 5' or at the 3' end of the nucleotide sequence. Typical algorithms will then identify oligomers of defined length that are unique to the gene, have a GC content within a range suitable for hybridization, and lack predicted secondary structure that may interfere with hybridization. In certain situations it may be appropriate to use pairs of oligonucleotides on a microarray. The

"pairs" will be identical, except for one nucleotide that preferably is located in the center of the sequence. The second oligonucleotide in the pair (mismatched by one) serves as a control. The number of oligonucleotide pairs may range from two to one million. The oligomers are synthesized at designated areas on a substrate using a light-directed chemical process. The substrate may be paper, nylon or other type of membrane, filter, chip, glass slide or any other suitable solid support.

In another aspect, an oligonucleotide may be synthesized on the surface of the substrate by using a chemical coupling procedure and an ink jet application apparatus, as described in PCT application W095/251116 (Baldeschweiler et al.) which is incorporated herein in its entirety by reference. In another aspect, a "gridded" array analogous to a dot (or slot) blot may be used to arrange and link cDNA fragments or oligonucleotides to the surface of a substrate using a vacuum system, thermal, UV, mechanical or chemical bonding procedures. An array, such as those described above, may be produced by hand or by using available devices (slot blot or dot blot apparatus), materials (any suitable solid support), and machines (including robotic instruments), and may contain 8, 24, 96, 384, 1536, 6144 or more oligonucleotides, or any other number between two and one million which lends itself to the efficient use of commercially available instrumentation.

In order to conduct sample analysis using a microarray, the RNA or DNA from a biological sample is made into hybridization probes. The mRNA is isolated, and cDNA is produced and used as a template to make antisense RNA (aRNA). The aRNA is amplified in the presence of fluorescent nucleotides, and labeled probes are incubated with the microarray so that the probe sequences hybridize to complementary oligonucleotides of the microarray. Incubation conditions are adjusted so that hybridization occurs with precise complementary matches or with various degrees of less complementarity. After removal of nonhybridized probes, a scanner is used to determine the levels and patterns of fluorescence. The scanned images are examined to determine degree of complementarity and the relative abundance of each oligonucleotide sequence on the microarray. The biological samples may be obtained from any bodily fluids (such as blood, urine, saliva, phlegm, gastric juices, etc.), cultured cells, biopsies, or other tissue preparations. A detection system may be used to measure the absence, presence, and amount of hybridization for all of the distinct sequences simultaneously. This data

may be used for large-scale correlation studies on the sequences, expression patterns, mutations, variants, or polymorphisms among samples.

Using such arrays, the present invention provides methods to identify the expression of one or more of the proteins/peptides of the present invention. In detail, such methods comprise incubating a test sample with one or more nucleic acid molecules and assaying for binding of the nucleic acid molecule with components within the test sample. Such assays will typically involve arrays comprising many genes, at least one of which is a gene of the present invention. Figure 3 provides information on SNPs that have been found in the gene encoding the Ras-like protein of the present invention. SNPs were identified at 25 different nucleotide positions. Some of these SNPs that are located outside the ORF and in introns may affect gene transcription.

Conditions for incubating a nucleic acid molecule with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the nucleic acid molecule used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification or array assay formats can readily be adapted to employ the novel fragments of the human genome disclosed herein. Examples of such assays can be found in Chard, T, *An Introduction to Radioimmunoassay and Related Techniques*, Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock, G. R. *et al.*, *Techniques in Immunocytochemistry*, Academic Press, Orlando, FL Vol. 1 (1982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., *Practice and Theory of Enzyme Immunoassays: Laboratory Techniques in Biochemistry and Molecular Biology*, Elsevier Science Publishers, Amsterdam, The Netherlands (1985).

The test samples of the present invention include cells, protein or membrane extracts of cells. The test sample used in the above-described method will vary based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing nucleic acid extracts or of cells are well known in the art and can be readily be adapted in order to obtain a sample that is compatible with the system utilized.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the assays of the present invention.

Specifically, the invention provides a compartmentalized kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the nucleic acid molecules that can bind to a fragment of the human genome disclosed herein; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of a bound nucleic acid. Preferred kits will include chips that are capable of detecting the expression of 10 or more, 100 or more, or 500 or more, 1000 or more, or all of the genes expressed in Human.

In detail, a compartmentalized kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers, strips of plastic, glass or paper, or arraying material such as silica. Such containers allows one to efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the nucleic acid probe, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, etc.), and containers which contain the reagents used to detect the bound probe. One skilled in the art will readily recognize that the previously unidentified Ras-like protein genes of the present invention can be routinely identified using the sequence information disclosed herein can be readily incorporated into one of the established kit formats which are well known in the art, particularly expression arrays.

Vectors/host cells

The invention also provides vectors containing the nucleic acid molecules described herein. The term "vector" refers to a vehicle, preferably a nucleic acid molecule, which can transport the nucleic acid molecules. When the vector is a nucleic acid molecule, the nucleic acid molecules are covalently linked to the vector nucleic acid. With this aspect of the invention, the vector includes a plasmid, single or double stranded phage, a single or double stranded RNA or DNA viral vector, or artificial chromosome, such as a BAC, PAC, YAC, OR MAC.

A vector can be maintained in the host cell as an extrachromosomal element where it replicates and produces additional copies of the nucleic acid molecules. Alternatively, the vector may integrate into the host cell genome and produce additional copies of the nucleic acid molecules when the host cell replicates.

The invention provides vectors for the maintenance (cloning vectors) or vectors for expression (expression vectors) of the nucleic acid molecules. The vectors can function in procaryotic or eukaryotic cells or in both (shuttle vectors).

Expression vectors contain cis-acting regulatory regions that are operably linked in the vector to the nucleic acid molecules such that transcription of the nucleic acid molecules is allowed in a host cell. The nucleic acid molecules can be introduced into the host cell with a separate nucleic acid molecule capable of affecting transcription. Thus, the second nucleic acid molecule may provide a trans-acting factor interacting with the cis-regulatory control region to allow transcription of the nucleic acid molecules from the vector.

Alternatively, a trans-acting factor may be supplied by the host cell. Finally, a trans-acting factor can be produced from the vector itself. It is understood, however, that in some embodiments, transcription and/or translation of the nucleic acid molecules can occur in a cell-free system.

The regulatory sequence to which the nucleic acid molecules described herein can be operably linked include promoters for directing mRNA transcription. These include, but are not limited to, the left promoter from bacteriophage λ , the lac, TRP, and TAC promoters from *E. coli*, the early and late promoters from SV40, the CMV immediate early promoter, the adenovirus early and late promoters, and retrovirus long-terminal repeats.

In addition to control regions that promote transcription, expression vectors may also include regions that modulate transcription, such as repressor binding sites and enhancers. Examples include the SV40 enhancer, the cytomegalovirus immediate early enhancer, polyoma enhancer, adenovirus enhancers, and retrovirus LTR enhancers.

In addition to containing sites for transcription initiation and control, expression vectors can also contain sequences necessary for transcription termination and, in the transcribed region a ribosome binding site for translation. Other regulatory control elements for expression include initiation and termination codons as well as polyadenylation signals. The person of ordinary skill in the art would be aware of the numerous regulatory sequences

that are useful in expression vectors. Such regulatory sequences are described, for example, in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*. 2nd. ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1989).

A variety of expression vectors can be used to express a nucleic acid molecule. Such vectors include chromosomal, episomal, and virus-derived vectors, for example vectors derived from bacterial plasmids, from bacteriophage, from yeast episomes, from yeast chromosomal elements, including yeast artificial chromosomes, from viruses such as baculoviruses, papovaviruses such as SV40, Vaccinia viruses, adenoviruses, poxviruses, pseudorabies viruses, and retroviruses. Vectors may also be derived from combinations of these sources such as those derived from plasmid and bacteriophage genetic elements, e.g. cosmids and phagemids. Appropriate cloning and expression vectors for prokaryotic and eukaryotic hosts are described in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*. 2nd. ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1989).

The regulatory sequence may provide constitutive expression in one or more host cells (i.e. tissue specific) or may provide for inducible expression in one or more cell types such as by temperature, nutrient additive, or exogenous factor such as a hormone or other ligand. A variety of vectors providing for constitutive and inducible expression in prokaryotic and eukaryotic hosts are well known to those of ordinary skill in the art.

The nucleic acid molecules can be inserted into the vector nucleic acid by well-known methodology. Generally, the DNA sequence that will ultimately be expressed is joined to an expression vector by cleaving the DNA sequence and the expression vector with one or more restriction enzymes and then ligating the fragments together. Procedures for restriction enzyme digestion and ligation are well known to those of ordinary skill in the art.

The vector containing the appropriate nucleic acid molecule can be introduced into an appropriate host cell for propagation or expression using well-known techniques. Bacterial cells include, but are not limited to, *E. coli*, *Streptomyces*, and *Salmonella typhimurium*. Eukaryotic cells include, but are not limited to, yeast, insect cells such as *Drosophila*, animal cells such as COS and CHO cells, and plant cells.

As described herein, it may be desirable to express the peptide as a fusion protein. Accordingly, the invention provides fusion vectors that allow for the production of the

peptides. Fusion vectors can increase the expression of a recombinant protein, increase the solubility of the recombinant protein, and aid in the purification of the protein by acting for example as a ligand for affinity purification. A proteolytic cleavage site may be introduced at the junction of the fusion moiety so that the desired peptide can ultimately be separated from the fusion moiety. Proteolytic enzymes include, but are not limited to, factor Xa, thrombin, and enteroRas-like protein. Typical fusion expression vectors include pGEX (Smith *et al.*, *Gene* 67:31-40 (1988)), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann *et al.*, *Gene* 69:301-315 (1988)) and pET 11d (Studier *et al.*, *Gene Expression Technology: Methods in Enzymology* 185:60-89 (1990)).

Recombinant protein expression can be maximized in a host bacteria by providing a genetic background wherein the host cell has an impaired capacity to proteolytically cleave the recombinant protein. (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Alternatively, the sequence of the nucleic acid molecule of interest can be altered to provide preferential codon usage for a specific host cell, for example *E. coli*. (Wada *et al.*, *Nucleic Acids Res.* 20:2111-2118 (1992)).

The nucleic acid molecules can also be expressed by expression vectors that are operative in yeast. Examples of vectors for expression in yeast e.g., *S. cerevisiae* include pYepSec1 (Baldari, *et al.*, *EMBO J.* 6:229-234 (1987)), pMFa (Kurjan *et al.*, *Cell* 30:933-943(1982)), pJRY88 (Schultz *et al.*, *Gene* 54:113-123 (1987)), and pYES2 (Invitrogen Corporation, San Diego, CA).

The nucleic acid molecules can also be expressed in insect cells using, for example, baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith *et al.*, *Mol. Cell Biol.* 3:2156-2165 (1983)) and the pVL series (Lucklow *et al.*, *Virology* 170:31-39 (1989)).

In certain embodiments of the invention, the nucleic acid molecules described herein are expressed in mammalian cells using mammalian expression vectors. Examples of

mammalian expression vectors include pCDM8 (Seed, B. *Nature* 329:840(1987)) and pMT2PC (Kaufman *et al.*, *EMBO J.* 6:187-195 (1987)).

The expression vectors listed herein are provided by way of example only of the well-known vectors available to those of ordinary skill in the art that would be useful to express the nucleic acid molecules. The person of ordinary skill in the art would be aware of other vectors suitable for maintenance, propagation, or expression of the nucleic acid molecules described herein. These are found for example in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual*. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

The invention also encompasses vectors in which the nucleic acid sequences described herein are cloned into the vector in reverse orientation, but operably linked to a regulatory sequence that permits transcription of antisense RNA. Thus, an antisense transcript can be produced to all, or to a portion, of the nucleic acid molecule sequences described herein, including both coding and non-coding regions. Expression of this antisense RNA is subject to each of the parameters described above in relation to expression of the sense RNA (regulatory sequences, constitutive or inducible expression, tissue-specific expression).

The invention also relates to recombinant host cells containing the vectors described herein. Host cells therefore include prokaryotic cells, lower eukaryotic cells such as yeast, other eukaryotic cells such as insect cells, and higher eukaryotic cells such as mammalian cells.

The recombinant host cells are prepared by introducing the vector constructs described herein into the cells by techniques readily available to the person of ordinary skill in the art. These include, but are not limited to, calcium phosphate transfection, DEAE-dextran-mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, lipofection, and other techniques such as those found in Sambrook, *et al.* (*Molecular Cloning: A Laboratory Manual*. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

Host cells can contain more than one vector. Thus, different nucleotide sequences can be introduced on different vectors of the same cell. Similarly, the nucleic acid molecules can be introduced either alone or with other nucleic acid molecules that are not

related to the nucleic acid molecules such as those providing trans-acting factors for expression vectors. When more than one vector is introduced into a cell, the vectors can be introduced independently, co-introduced, or joined to the nucleic acid molecule vector.

In the case of bacteriophage and viral vectors, these can be introduced into cells as packaged or encapsulated virus by standard procedures for infection and transduction. Viral vectors can be replication-competent or replication-defective. In the case in which viral replication is defective, replication will occur in host cells providing functions that complement the defects.

Vectors generally include selectable markers that enable the selection of the subpopulation of cells that contain the recombinant vector constructs. The marker can be contained in the same vector that contains the nucleic acid molecules described herein or may be on a separate vector. Markers include tetracycline or ampicillin-resistance genes for prokaryotic host cells and dihydrofolate reductase or neomycin resistance for eukaryotic host cells. However, any marker that provides selection for a phenotypic trait will be effective.

While the mature proteins can be produced in bacteria, yeast, mammalian cells, and other cells under the control of the appropriate regulatory sequences, cell-free transcription and translation systems can also be used to produce these proteins using RNA derived from the DNA constructs described herein.

Where secretion of the peptide is desired, which is difficult to achieve with multi-transmembrane domain containing proteins such as kinases, appropriate secretion signals are incorporated into the vector. The signal sequence can be endogenous to the peptides or heterologous to these peptides.

Where the peptide is not secreted into the medium, which is typically the case with kinases, the protein can be isolated from the host cell by standard disruption procedures, including freeze thaw, sonication, mechanical disruption, use of lysing agents and the like. The peptide can then be recovered and purified by well-known purification methods including ammonium sulfate precipitation, acid extraction, anion or cationic exchange chromatography, phosphocellulose chromatography, hydrophobic-interaction chromatography, affinity chromatography, hydroxylapatite chromatography, lectin chromatography, or high performance liquid chromatography.

It is also understood that depending upon the host cell in recombinant production of the peptides described herein, the peptides can have various glycosylation patterns, depending upon the cell, or maybe non-glycosylated as when produced in bacteria. In addition, the peptides may include an initial modified methionine in some cases as a result of a host-mediated process.

Uses of vectors and host cells

The recombinant host cells expressing the peptides described herein have a variety of uses. First, the cells are useful for producing a Ras-like protein polypeptide that can be further purified to produce desired amounts of Ras-like protein or fragments. Thus, host cells containing expression vectors are useful for peptide production.

Host cells are also useful for conducting cell-based assays involving the Ras-like protein or Ras-like protein fragments. Thus, a recombinant host cell expressing a native Ras-like protein is useful for assaying compounds that stimulate or inhibit Ras-like protein function.

Host cells are also useful for identifying Ras-like protein mutants in which these functions are affected. If the mutants naturally occur and give rise to a pathology, host cells containing the mutations are useful to assay compounds that have a desired effect on the mutant Ras-like protein (for example, stimulating or inhibiting function) which may not be indicated by their effect on the native Ras-like protein.

Genetically engineered host cells can be further used to produce non-human transgenic animals. A transgenic animal is preferably a mammal, for example a rodent, such as a rat or mouse, in which one or more of the cells of the animal include a transgene. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal in one or more cell types or tissues of the transgenic animal. These animals are useful for studying the function of a Ras-like protein and identifying and evaluating modulators of Ras-like protein activity. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, and amphibians.

A transgenic animal can be produced by introducing nucleic acid into the male pronuclei of a fertilized oocyte, e.g., by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. Any of the Ras-like protein

nucleotide sequences can be introduced as a transgene into the genome of a non-human animal, such as a mouse.

Any of the regulatory or other sequences useful in expression vectors can form part of the transgenic sequence. This includes intronic sequences and polyadenylation signals, if not already included. A tissue-specific regulatory sequence(s) can be operably linked to the transgene to direct expression of the Ras-like protein to particular cells.

Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, both by Leder *et al.*, U.S. Patent No. 4,873,191 by Wagner *et al.* and in Hogan, B., *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the transgene in its genome and/or expression of transgenic mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene can further be bred to other transgenic animals carrying other transgenes. A transgenic animal also includes animals in which the entire animal or tissues in the animal have been produced using the homologously recombinant host cells described herein.

In another embodiment, transgenic non-human animals can be produced which contain selected systems which allow for regulated expression of the transgene. One example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For a description of the *cre/loxP* recombinase system, see, e.g., Lakso *et al. PNAS* 89:6232-6236 (1992). Another example of a recombinase system is the FLP recombinase system of *S. cerevisiae* (O'Gorman *et al. Science* 251:1351-1355 (1991)). If a *cre/loxP* recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the *Cre* recombinase and a selected protein is required. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, I. *et al. Nature* 385:810-813 (1997) and PCT International Publication Nos. WO 97/07668 and WO 97/07669. In brief, a cell, e.g., a somatic cell, from the transgenic animal can be isolated and induced to exit the growth cycle and enter G₀ phase. The quiescent cell can then be fused, e.g., through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyst and then transferred to pseudopregnant female foster animal. The offspring born of this female foster animal will be a clone of the animal from which the cell, e.g., the somatic cell, is isolated.

Transgenic animals containing recombinant cells that express the peptides described herein are useful to conduct the assays described herein in an *in vivo* context. Accordingly, the various physiological factors that are present *in vivo* and that could effect ligand binding, Ras-like protein activation, and signal transduction, may not be evident from *in vitro* cell-free or cell-based assays. Accordingly, it is useful to provide non-human transgenic animals to assay *in vivo* Ras-like protein function, including ligand interaction, the effect of specific mutant Ras-like proteins on Ras-like protein function and ligand interaction, and the effect of chimeric Ras-like proteins. It is also possible to assess the effect of null mutations, which is mutations that substantially or completely eliminate one or more Ras-like protein functions.

All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the above-described modes for carrying out the invention, which are obvious to those skilled in the field of molecular biology or related fields, are intended to be within the scope of the following claims.

SEQUENCE LISTING

<110> MERKULOV, Gennady et al.

<120> ISOLATED HUMAN RAS-LIKE PROTEINS,
NUCLEIC ACID MOLECULES ENCODING THESE HUMAN RAS-LIKE
PROTEINS, AND USES THEREOF

<130> CL001196

<160> 4

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 1405

<212> DNA

<213> Human

<400> 1

```

aagcgaatagc tgagtgcggc ggctgctgat tgtgttctag gggacggagt aggggaagac 60
gtttgtctctc ccggaacagc ctatctcatt cctttctttc gattaccgtt ggccgcggaga 120
gtcaggggcgg cggtgcggc agcaaggcgg gcggtggcgg cgccggcagc tgcagtgcaga 180
tgtccagcat gaatcccgaa tatgattatt tattcaagtt acttctgatt ggcgactcag 240
gggttgaaaa gtcttgccct cttcttaggt ttgcagatga tacatatata gaaagctaca 300
tcagcacaat tgggtgtgat ttcaaaataa gaactataga gttagacggg aaaacaatca 360
agcttcaaat agagtccttc aataatgtta aacagtggct gcaggaaata gatcgttatg 420
ccagtgaataa tgtcaacaaa ttgttggtag ggaacaaatg tgatctgacc acaaagaaaag 480
tagtagacta cacaacagcg aaggaatttg ctgattccct tgggaattccg tttttggaaa 540
ccagtgtctaa gaatgcaacg aatgtagaac agtctttcat gacgatggca gctgagatta 600
aaaagcgaat gggtcccggg gcaacagctg gtgtgtctga gaagtccaat gttaaaattc 660
agagcactcc agtcaagcag tcagggtggag gttgctgcta aaatttgctt ccatcctttt 720
ctcacagcaa tgaatttgca atctgaacc aagtgaaaaa acaaaattgc ctgaattgta 780
ctgtatgtag ctgcactaca acagattctt accgtctcca caaaggtcag agattgtaaa 840
tggtcaatac tgactttttt tttattccct tgactcaaga cagctaactt cattttcaga 900
actgttttaa acctttgtgt gctggtttat aaaataatgt gtgtaattct tgttgctttc 960
ctgataccag actgtttccc gtggttggtt agaataatgt ttgttttgat gtttatattg 1020
gcatgtttag atgtcaggtt tagtcttctg aagatgaagt tcagccattt tgtatcaaac 1080
agcacaagca gtgtctgtca ctttccatgc ataaagttaa gtgagatgtt atatgtaaga 1140
tctgatttgc tagttcttcc ttgtagagtt ataaatggaa agattacact atctgattaa 1200
tagtttcttc atactctgca tataatttgt ggctgcagaa tattgttaatt tgttgcacac 1260
tatgtaacaa aacaactgaa gatattgtta ataaatattg tacttattgg aagtaaaaaa 1320
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1380
aaaaaaaaaa aaaaaaaaaa aaaaaa

```

<210> 2

<211> 173

<212> PRT

<213> Human

<400> 2

```

Met Ser Ser Met Asn Pro Glu Tyr Asp Tyr Leu Phe Lys Leu Leu Leu
  1           5           10           15
Ile Gly Asp Ser Gly Val Gly Lys Ser Cys Leu Leu Leu Arg Phe Ala
          20           25           30
Asp Asp Thr Tyr Thr Glu Ser Tyr Ile Ser Thr Ile Gly Val Asp Phe
          35           40           45

```

Lys Ile Arg Thr Ile Glu Leu Asp Gly Lys Thr Ile Lys Leu Gln Ile
 50 55 60
 Glu Ser Phe Asn Asn Val Lys Gln Trp Leu Gln Glu Ile Asp Arg Tyr
 65 70 75 80
 Ala Ser Glu Asn Val Asn Lys Leu Leu Val Gly Asn Lys Cys Asp Leu
 85 90 95
 Thr Thr Lys Lys Val Val Asp Tyr Thr Thr Ala Lys Glu Phe Ala Asp
 100 105 110
 Ser Leu Gly Ile Pro Phe Leu Glu Thr Ser Ala Lys Asn Ala Thr Asn
 115 120 125
 Val Glu Gln Ser Phe Met Thr Met Ala Ala Glu Ile Lys Lys Arg Met
 130 135 140
 Gly Pro Gly Ala Thr Ala Gly Gly Ala Glu Lys Ser Asn Val Lys Ile
 145 150 155 160
 Gln Ser Thr Pro Val Lys Gln Ser Gly Gly Gly Cys Cys
 165 170

<210> 3
 <211> 46050
 <212> DNA
 <213> Human

<220>
 <221> misc_feature
 <222> (1)...(46050)
 <223> n = A,T,C or G

<400> 3
 ttttggtgtgt gtgtgtgtgt gtgtgtgtgt gtgcctttac tagtgactca ggtcacagtt 60
 ttctgagatt ttttttctcc cctcaagaca gaatcttgct ctgtcgccca ggctggagtg 120
 cagtggcctc tcggcccaact gtagecctccg cctcccgggt tcaagcaatt ttcctgcctc 180
 agcctcccga gtagctggga ttacaggcac gcgccaccat gcctggctaa tttttgtatt 240
 tttagtagag acagtgtttc accatgttggt ccaggctggt cttgaattcc tgacctcgtg 300
 atctgtccgt tttggcctct caaatcctct agattacagg catgagccac cgagcctggc 360
 cagttttctg agtttttatt tgaaatcaaa ataagctttt tttttttttt taatgggctt 420
 tagagtcag ggtaacgaac acttttttgt gcctattact gaaccattca gggatttcct 480
 ggggtggtga ccgtgttcat ttcagaaacc aacatgttca tttcagaaac caaactcggg 540
 taacttttga taagtctatc aactaaggcc catggcagaa tttgagggt aaggggtgta 600
 attagtgtat gggtagaaat aagtgccttc tttctatatt ttggcgttgt aggaatttaa 660
 agtgattctg cagtaagtct caggagacaa ttttcttagt tcttagaagt tggaaagataa 720
 actttggaca atgtattaca ctatgccctt tgtaattaaa taactcaaga taatgtgtta 780
 aagtttagcg gagatttaaa ttcctgagct gattaaagag agctgttaag gccatagggt 840
 ttttaaaaat gagttaatat tactcccaga aattgtaggc actatatagt gatgaattgc 900
 atatttttat tgcttattat tttccagtct tgcagaatgg ctcagggtta gtagcaacta 960
 aaagataata cattacaatt caacctgaag gccgggacga aggtagggaat tggatttttag 1020
 gctggctctg ggctgtgtcc ctcccatcca tgggatgtgg agccattgaa ggttgtgggg 1080
 tcacgatgca ggtgctgtct cagaaagata catccgactg tgtgtgcaaa tgggctgggg 1140
 cgagagaagag agagagaggt agagtccatt tggagactac tgcaatagcc aggctgacga 1200
 gttaagagcg gggcacagta agaatgggaa gaaatctaag aagaaaatgg tagtgcgcgg 1260
 ggccaacaat ggacgatgac cgaaccagag tggggatggg tgagtgcga gaagaaccgc 1320
 tccgtgccgt ccagggagcc ccttgacttc cttctgttc ttagagcgga cgtcctccta 1380
 ccagccccc accagcgcca ccagggtggc gcaagcctca agctggtcag gtcagcaaca 1440
 gccgcaacgg aggcaggagc cgacacgctc gtaccccggc cccctccccg ccccgccacc 1500
 cccggcagtc cctccggttt gaccactccc cccggtcctt tgccctcccc gacccccagc 1560
 ctccgtcggc cgccggcacc accctccgcc cctctccgcc ccctccccg tggggcgctg 1620
 actcgcccg ctgccacgtc tactgatga catcactagg gcagctcggc cttagccaat 1680

ccgccagggg gagtccgagc gaagtcctag ccagcgagtc agaggggagg ggagcagggg 1740
 ggggcccagg gtggggagggt gagggagtggt ggaatggggc gggcgacaac cttcaggtg 1800
 cgcatgcccc agaggcgagg cgcttgccgg gaagctgagt cctggccttg cgtcgactg 1860
 tctgtcctca gctcgcgtag ccgcgctcgc gactcccttt cccggcatgc caggcgggtg 1920
 ggccgccttc tggggccgtgt aaaggccctt cgggtctaagg cttccctatt tctgtgttcg 1980
 ccggcgggcca ttttgggtgg aagcgatagc tgagtggcgg cggctgctga ttgtgttcta 2040
 ggggacggag taggggaaga cgtttgctct cccggaacag cctatctcat tcccttcttt 2100
 cgattacccg tggcgcgagg agtcaggcg gcggtcgagg cagcaagggc ggcggtggcg 2160
 gcggcggcag ctgcatgac atgtccagca tgaatcccga atagttagtt caggagagca 2220
 ccggtcgggt ggggtccgtgg gccagcttgg gggatcttaa aggggtcgag gagggttggg 2280
 gcagaagtcg gggcatcggg tggggtgagg aggggtgat gggtcaggag aggctggcgg 2340
 ccgggagtcg ggccctattg tctgacggg agggcgggcc gcgcggggga ggggtcgggc 2400
 cggaggggtg agccgcccgg gcctggaccg ggtcaggtta gagggcctga ctgcgggcg 2460
 ggtgctgagg aagcctgccc aggggcctgg ggcggtgta aggggtatct tctctcgag 2520
 gcagtgactt ttgaaggagg acttgtctct aaggggagg gatgggtgg gagagccctt 2580
 ctagagggca ctgtcagacc ctgcccgcg actctcgga gctgtcagga tcttcgggtg 2640
 agaaaccagc ttactttgta aatcctgagc ttgttgggtc tctctccttc catcctcccc 2700
 gccaggtttc agtgaatatg gatgcttttc gggactgcgt gggattgagg ggaatgagta 2760
 gatggtgaga agcaactgaa catttattag ttctcttttt gagttgtgtc ttggaggagt 2820
 tgtttaagag ctgcccgggt ccattgccct cctataaaaa cctgggcatt tgtgagaatt 2880
 ttgttttttt tttttttaa gaggacacct aagtcatttt gtcttctgtg ggtcaaggga 2940
 aaaaaaaaaa actaaagcca agaatgtct ttttgatact cgcagattaa aggaagcttg 3000
 ctgtcaagtt gaaagagaaa cgaacgggac ctatgataga tctgtatgta ggttttggat 3060
 tacttgcttg gatgcttgca gatagggaat gaggttccat gacgtgtcat gaaaagttaa 3120
 tgcatttctt tttcttgctt actcaagaag tcaccacagc agatgtgaca cacctggcac 3180
 ctttccctgg aactgggtgtt cacttccctt gggtagagtt tgttgggctc tctcaatgg 3240
 ccctttaaaa atttctctta cagtttacat gcatgtaaag taatgaataa ttggaagaga 3300
 ccgaatttgt attccttttc agtgtcaaag gcctttgagg gatgggggaa aatcagtatt 3360
 tgttgtaaaa gttgagttta tttgctggtt tggtaatta ctgctagaca ttttccccta 3420
 aaaggtccac ccaccagttt agctgactgt catatgtgtg tcacatggct cttgcaaaat 3480
 gcttacaagt tttgtaatat tgtggcttga agctgaaatc ttttgacta aacagaaacc 3540
 gtagtatttt attagaattt catgctttag aagttgagg tagtggttct gtagtgacat 3600
 ttgtgtgtt gacagtttaa aaaaattttt ttttcaagg ctccaaggac aaagttggtt 3660
 ttgcacagtt gaacggagg gaacttgagg ttcttaattt agtagttttc ttggttaaca 3720
 taaagaacat ggatttactg ctttatcgag gtttatagac ctctactgtt caggaaattt 3780
 tctgaatttg ctatatatat gtttattagt gtaaataaat cttcaagatt agttgagaac 3840
 tttgacaagt tactcagcct ctgaattttt tttccctttt gtaaaatagg ataattggag 3900
 tcattattcc tgtcagggtg gtgggtgaaat tcaaagtgt ataaaaaat ttgaaaaact 3960
 gtgtgagcat tcttcagggt gtagtcatca ttttcatgaa aggcattcta ttagtaccag 4020
 gatttaggaa tataatcctt gcgcttaaga agtttagata taggccaggc gcgggtggctc 4080
 acctcagtaa tcccagcact ttgggaggcc gaggcggggc gatcccgagg tcaggagatc 4140
 gagaccatcc tcggtaacac ggtgaaaccc cgtctctact aaaaatgcaa aaaaattagc 4200
 cgggctggtt ggtgggcacc tgtagtccca gctactcgag aggctgaggc aggagaatgg 4260
 cgtgatcccg ggaggtggag ctgcatgta accaagatct ggccactgca ctccagcctg 4320
 gacgacagag caagactccg tctcaaaaaa aaaattattt attgttttga gacggagttt 4380
 caatcttggt gccagggtg gagtgcattg gcgcaaatct cctctcaccg ccacctccgc 4440
 ctctgggtt caagtattc tctgcctca gattcccag aagttgggat tacaggcatg 4500
 tgccaccact ccgggctaatt tttgtatttt tggtagagac ggggtttctc catgttggct 4560
 aggtgtgtct caaactcccg aagtatccg ccgcctcag cttcccaaag tgttgggtt 4620
 acaggcgtga gccaccgcgc ccggcagaaa tagattttat acatgtcaaa taccagtaga 4680
 tatagcaaat tccagatgtg tggcatggat gagagcaaca agatttcagg gggatgggtg 4740
 gttgtggtt gctatctggg ttttgaaga ctttatagaa gagagacctg aaagggattt 4800
 atcagcaatt agatttgagg gaacagagg agtgactagg aattttcaag ggggagaaga 4860
 aggaggaatg gtcataaat gacaaggaca gtaataagta aatacgggt caaatcatcc 4920
 tttcttttga agactaatga cctcaaagg atcaaaccca gaaacagttt ttatatattt 4980
 tctgggatca aatacatgg tatctggcct actatattt tattctagac tgttagtaga 5040
 aataatacag gaatttgaga aaacctttgc aaaagtgtta gtgaaaatta cttaggggtg 5100

gaggaagtga	gggatatttt	attaggggag	gtcacaaagg	cagtgaagca	tcagattttt	5160
agtaaatctga	cttaagcagt	ttctttttgt	tttaaatgaag	cttgttatct	ttataaaaagt	5220
aattagagaa	aatttgga	ataaaggaaa	gaaagaaaag	ttcttttagtg	ttttatcacg	5280
caaatacaag	ctcattcggt	tttaacatct	tggtccaaac	tccaaagtct	tgctttctct	5340
tcaattaaaa	ctttaatggg	tggatgcttt	tcctgcttcc	agtatgttat	cttaataaact	5400
aacaatggta	tattagctaa	tgtttacaaa	tgtactccag	atgttcctta	agttactttg	5460
gtttatcatt	accaatttat	attgtttcct	ttagaaattt	ataatctttg	ttaatgggtt	5520
ctgctaaatt	tggtagtga	aatgggatct	tgagaaaaa	gattctgaag	caacagaatt	5580
tttagattta	tattggttta	cataagagtt	ggtagctgta	ttactttttt	tgtttggttt	5640
gttttttttt	tgagacggaa	tcttgctctg	tcgccaggc	cttggcctcc	caaagtgttg	5700
ggattacagg	cgtgagccac	tgtgcctggc	tgtttggtt	ttttttgtt	tttgttttct	5760
tttctttttc	tttttttcga	gatggagtct	cactctgtca	cccaggctgg	agtgcagtgg	5820
cgcgatcttg	gctcactgca	atctctgcct	cctgggttca	agcgattttc	ctgccttggt	5880
ctctgagta	gctgggatta	caggcatttg	ccaccataac	cagctaattt	ttgtatagag	5940
taccagcca	tctctaattg	tgatcaggct	gaagcagggt	gatcacctaa	ggtcaggagt	6000
tcaagaccag	cctggccaat	atggcaaaac	cctatctcta	ctaatacaga	aaattgtctg	6060
ggtgtgttgg	ctggcgctg	taatcccagc	tactcgggag	gctgaggcag	gacaatctct	6120
tgaacctcgg	aggtggaggt	tgcagtgaac	cgagatcaca	ccattgcact	ccagcctggg	6180
caacagagca	agacttgtct	caaaaaaaaa	aaaaaaaaa	aaaaaaaggc	aattgaaagt	6240
gtaatctgaa	cagttaaaaa	agtagataga	aagggttaaa	gctttttttt	gaggatctga	6300
agaaaaatgt	ggattttttt	tgagctacgt	tttgaagcag	gcagtgatta	tttcagcaca	6360
ttaagaaatg	cttaacatgg	ccaggcgagc	tggtcacgc	ctgtaattct	cagcactttg	6420
ggaggccgag	tgggcggtat	catttgaggt	cagaccagc	ctggccaaca	tgatgagaca	6480
ctgcctctac	taaaaataca	aaaattagct	gggtgtggtg	gtgcacgcct	gtaattccag	6540
ctactcagga	acctgaggca	ggagagtcac	ttgaacctgg	gaggcgagg	ctgcagtga	6600
tccagatcat	gccactgcac	tccagcctga	gggacagagt	gagactcctc	aaaaaaaaaa	6660
aaaaaaaaaag	aaagaaatac	ttaacattat	tctcgtgatt	attctcataa	catttttcat	6720
aatccactgg	cttcagtggt	atttttttag	tgtcaagaaa	ataattttga	ttgggttcac	6780
tttaaggaat	gtgttaagaa	taaagcatgt	ctacctgtct	tcagtatacc	agctaactat	6840
agtaggaaga	aatatagtag	tctacttaga	tcaactataa	ttctttaatg	cagaaaaagt	6900
ttaaagtatt	taccttattt	ttagcccca	tccccttaag	tatatcatgg	ctccagaatc	6960
tctgaaaatg	ttatcagttc	ttcagacttt	gctcttcttt	catgttatac	tcaagaaaca	7020
tttgaccttt	tttttttttt	ttttgcttgc	attgtgtttc	aaataatttt	taacaaaact	7080
taagtgtttg	aaagtgaag	cagggtgtct	ttgtgacttt	tggtgggtgg	tgaaaaaact	7140
cagaaaagtt	taaagaagaa	agataactag	tattctcatt	gtccagaata	tgatttttta	7200
aatgtctata	gaatatcacc	atctgtaatt	cttcgggtaa	tttaagtatt	cagtagttgt	7260
ataaaacctt	taaaatatat	atattgagaa	ttttgtgtga	atgagatgat	gagataatct	7320
tgtaaagtca	tttaagata	agaactgagg	cctggcacag	tggtcatgc	ctataatcac	7380
agcacttttg	gaggcccgag	cggtagatca	cctgaggtca	ggagtttgag	accagcctgg	7440
ccaacatggc	aaaaccctgt	ctctactaag	catagaaaa	tttaattgggt	gtggctcgtc	7500
ctgcgtgtag	tcccagctgc	ttgggaagct	gaggcgggag	aatctcttga	accctggagg	7560
tgggcattgc	agtgaagctga	gattgcgcca	ctgcactcca	gcctgggcga	cagagcaaga	7620
ctctgtctca	aaataaagta	aaataaaatg	aagataacaa	ctgaaatttc	acattaaaaa	7680
tttttttgta	gcgactgtgc	ctcctatggt	gtgcaggctg	gtctcaaact	cctggcctca	7740
agcgatcctt	ccaaagcact	gggtgggcca	ccatgtccag	cctgaaattt	tgcatataaa	7800
aatttccgc	ttttggcttg	gcgaggtgtc	tcacgcctgt	aatagcagtt	tgaggaggcg	7860
aggcaggcag	atcacttgag	gtcagttcta	gaccggcctg	gccaatgtgg	tgaaaccctg	7920
cctctactaa	aaacaccaa	ttagctaggc	gtggtggtgt	gcgctttag	tccaagcta	7980
ctgaggaggc	tgagacaaga	gaatcgcttg	aatctgggaa	aaagagggtg	ccgtgagcca	8040
agattggcca	ctgcactcca	gcctgggtga	cagagtgaga	ttctgtctca	aaaaataaaa	8100
aaataaaaaat	ttcccccttt	aatcaaatga	agttaaaatg	agggatgtta	gacagttttt	8160
aaccatcaaa	tatttttagtt	tagttttttt	tttttaacgt	tgtcttaaa	atggaagtgc	8220
ttcaaaatca	aatcttctct	gccagttctc	tacttggttt	cttttttttt	ctttttgaga	8280
tagagtctca	ctttgtcact	ggagtgcgtt	ggcgtgatct	cggctcactg	caacctccgc	8340
cttcagggtt	taagtgattc	ttccacctca	gcctctcaag	tagctgggag	tacagggtgtg	8400
tgccaccaca	cccggtcaat	ttttgtagtt	ttagtagaga	cagggtttca	ctatgttggtc	8460
caggctggcc	tcaaactcct	gacctcgtga	tccaccacc	tcagccaaat	tgctgggatt	8520

acttggtgga gccacgcgcc tggcttctac ttggctttta aagggaattt tgctttctga 8580
 gtaattttat ttctcaggta tcttggtcctt ttaattctg gaagcaatct taataattta 8640
 tgtatgtgcc ctgtaatccc agcactttgg gaggccgagg tgggcgaatc acgaggtcag 8700
 gagatcgaga ccacctctggc taacacgggtg aaaccccatc tactaaaaat acaaaaaatt 8760
 agctgggcgt ggtggcaggc gcctgtagtc ccagctactt nnnnnnnnnn nnnnnnnnnn 8820
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 8880
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 8940
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 9000
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 9060
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 9120
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 9180
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 9240
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 9300
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 9360
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 9420
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 9480
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 9540
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 9600
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 9660
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 9720
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 9780
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 9840
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 9900
 aacctctgtc tcccggttc cagcatttct tctgcctcag cctcctgagt aactgggact 9960
 acaggcgtcc accaccacgg ccagctaatt tttatattag tagagatggg gtttcacat 10020
 gttggccagg ctggtctcca actcctgacc tcaggtgata cgctgcctt ggtctccaa 10080
 agtgctagga ttacaggcgt gagccactac gtttggtgc ttatcagctt tttaccactt 10140
 tgtcgccact acatttttga attttctttt gagaattagg caaatgccc agactcccc 10200
 ccggccccc ctttagagg agaggggagc aattagacta ttctttgtt tccctataga 10260
 aggtggggct gagattactg ctttgatata tgggaatgaa tttagggaag aaaatttagg 10320
 tcttgccctt tcttggaac caccctggga gtgtgcaga ttattaatag ggtaatggtg 10380
 gaatgatatt caggggaaaa atggtcctga ggagccagag aactaagtgt tagtttggtg 10440
 gctgactgaa acatgtgaga gatagggtac agaagaagta ggaaatagtt ttccttggtg 10500
 cttctgtgac aggttggtc aattggctgg aacaccctac actgctttat taaatccaag 10560
 gttgtgata gttccagtta agtttactgt gttctatgct tgtagatttc ctaattagga 10620
 caagtagtgt taaatatgca tgcctttatt cacaagagg accattcttt tggaaacatc 10680
 actttttaat aatactagg gctatttagc acttactcgg tgccagccac gtggctatgg 10740
 ttttttttt ttttttttt cgagacatga tctagctctg tctcccaggc tggagtgggt 10800
 gtagcacagt catggctcac tgcagtctca acctcctgta ctctagtgat cctcctgtct 10860
 cagcctcctg agtaactggc acctgcctg gctaattttt ttttaagagat gagatgtcgc 10920
 tatgttgct atgctggct cgaacacctg ggctcaagt atcctccccg cctgagcctc 10980
 tcaaagtgtt gggattacag gtgtgaccca cctcacttgg ccactctatg tctttacata 11040
 gggcattttg tgcagtctgc atctcaaact agtgatcttc aacagtgaat ctcagtgaat 11100
 tatgtaattc atgttttcca agaacaatga tggatttaat ttctctgaat gtatttcctt 11160
 tgtataataa tagtacttaa gtggaattac tctttgtcct ttctactctc cttatagata 11220
 ttttctggta tcttgatttg ggactgttac atttaaccca tttatggctg tgtagccata 11280
 ctcacgttac atttgatgca tctgctcctt ttgtgtctat atactcatat aacattttgc 11340
 ataaagtatt aggcagttca caccaaggct gttcatgaac ctcagattaa gaatacttga 11400
 tttaggagat tgaaaaacaga aaagagaatg ttaactatca ttatcaatat taaaatgtga 11460
 aaatctgaga gtgacaaagc ttagctttta atctggtatc ccaaactcat ttgagttttt 11520
 ttttttttt ttttttttt gagacaaggt gtcgctttgt ccccaggct ggagtgtagt 11580
 ggtgtgatct tggtcactg caacctccac ctcccagggt caagtgattc tcctgcctca 11640
 gcctctgaag ttgctgggat tacaggtctg gccaccacgc ccagctaatt tttgtattt 11700
 atagtaaaga cggagtttca ccttattggc caggctggct tcaaaactct gatcttgtga 11760
 tcctcccgc tcggcctccc aaagtgcgtg gattacagg gtgagccact gttcccggcc 11820
 taatttgagt tttaaaatgt ggagttaaag atgttagtct taaagtgggt tagatgaaat 11880
 ttataaaaat agtcaaatag ctaaatttat aaaaggccat ttgaaacaat tttgtgaaat 11940

atataatgtg	gataaattatg	tagtgcttta	tgtgtagatt	ggtgggttagc	atctgcctga	12000
tgaagagcag	ttggatttct	tacttactaa	agctagttaa	atctgaactc	caaatttaggc	12060
atcttcacca	ggcttttttg	agccgagcta	acttactctc	ttttttatnt	ttattttttta	12120
attaattaat	tttttttttt	tttttttttt	tttggtagag	acaggatctc	cccatgttac	12180
ccaggcttgt	ctctggctcc	ttggctcaag	cagtcctcct	accttagcct	cccaaagtgc	12240
taggattaca	gctgtgagcc	actgcgccag	gctgagctta	ttctctacta	acacaagtgt	12300
tctaatttaa	tttaagcagt	gaatcacact	tttctttgta	tttggtcagg	ttctgggtgc	12360
tagtttata	atgatttgat	tcattctgat	agggtttttt	tggttttttt	tggtttttgt	12420
tttttgtttt	ttttgagaca	gagctctagct	ctgtcgccca	ggctggagtg	tggtggctcg	12480
atttcgggtc	attgcaactt	ctgcctccca	cccaggctgg	agtgcagtg	ctcgatttcg	12540
ggtcattgca	acctctgcct	cccaggttca	agcgattctc	ctgcctcagc	ctcctgagta	12600
gctgggatta	caagcaccca	ccaccatgcc	cggctaattt	tgtgtatttt	tagtagagac	12660
tgggtttcac	catgttgacc	acgctgggtc	cgaactcctg	acctcagtg	atctgcctgc	12720
cttggcctcc	caaagtgtct	ggattacagg	tgtgagccat	cacaccaggc	ctcaagaact	12780
ttttattttt	gagacaggtt	ctcactctgt	cacccaggct	ggagtacagt	ggtgagatca	12840
tggettactg	cagcctggac	ttcccaggct	ctggtgatcc	tcccatctca	gcccctggag	12900
taattaggaa	tatagacaca	cacccatgcc	tggcagtttt	tgtatttttt	ttcttttttc	12960
tctttttttg	tagagactgg	gtttcacatg	ttgtatcagg	ctgggttttg	actcctgagc	13020
tcaagcaatc	ctcactcttt	gacctcccaa	cgtgctggga	ttacaggcat	gagccactgt	13080
acctggcctt	ttctacatta	aaaaactttt	attaaaaaac	ccaaatcttc	cttgtggttg	13140
tatatacata	tatacatagg	tacacacatg	gagaatttta	ccttgaggga	aggcttggta	13200
aagaaaatag	ccctttgggc	cgggtgcggt	ggctgacgcc	tgtagtctta	gcactttggg	13260
aggctgaggt	gggcggattg	cctgagctca	ggagttcaag	accagcctgg	gcaacacagt	13320
gaaaccctgt	ctctactaaa	atacaaaaaa	tcagctgggt	gtggcagcat	gtgcctgttag	13380
tcccagctac	ttgggagcct	gaggcaggag	aactgcttga	accggggagg	cagaggttgc	13440
agtgagccga	gattgtgcta	ctgcacttca	gcctgcgcga	cagagcaaaa	ctctgtctca	13500
aaaaaacaaa	caaacaaaca	aaaaaggaaa	atagcctttc	tctatcatca	gagtatatta	13560
agagttgagt	ttttttttct	gttttttaaa	atttttgttg	tttattttta	attacaaaaac	13620
atggactctg	cttacaaatt	aagaaaatga	ctcatgttca	aacaagcata	atcaatataa	13680
cagttaatac	aagttaaata	ttgtaatatg	tttacggaat	agcatggcaa	aatagtgcga	13740
aagatttggg	gaaggggcct	ataatttctg	ttaacagaaa	gttttagtta	tggttgatcca	13800
actggagagg	aacagagctc	ccagaaggac	tccagaacac	ttgatgcttg	tctgagtggg	13860
gtcagcagca	ctgagttccc	accagccaga	aagtttgtgt	gtgtacatta	tttcccttaa	13920
ctgccacaat	aatcccata	agaaaatgcc	ctagttttac	aaacaaggaa	acagaggcag	13980
agaagagtta	aatgacttgc	ccaagggcat	tcaaagtaag	caactgaatt	ggaattttta	14040
ctcaaaggct	tggatgtccc	actacaacaa	ataggctgtt	tctgctttac	tacatgtgct	14100
tacttctaag	aatttaacat	tttaggctgg	tttgtgtggc	tactcctgt	aatctcagca	14160
ctttcgagg	ctgagtgagg	taaatcactt	gaagctcagga	gtttgagacc	aacctgggca	14220
acatggtaaa	acctcatctc	taccaaaaaa	aaaaaaaaaa	ctagctggac	gtggtggcac	14280
gcgcctgtgg	tcccagctac	tcaggaggct	gaagtaggag	gatcgtttga	gcctgggagg	14340
tggaggttgc	agttagccca	cattgcatca	ctgcactcta	gcctaggtga	cagagtgaga	14400
gcctatctca	cacacaaaaa	aaagaattta	aaattttagt	caagtaatta	ggcactaaca	14460
ttttgtggtc	agttacttta	cgaattcatg	gttggaggcc	tgatgtggtg	gctcatgcct	14520
gtaatcccag	cactttggga	ggctgaggca	ggaggattgc	ttaaggccaa	gagttcaaat	14580
cagcctgagc	aacctagtaa	gatccccctt	ctgcacaaaa	tttaaaaatt	agctgggcat	14640
ggtagtgtgc	acctgtagtc	ccaaccactt	gggaggctga	ggtgggagga	ttgcctgagg	14700
ccaggagttt	gagacctggg	cagcatatga	agaccctgtc	tctaaaaaac	taaaaataaa	14760
aaatagccag	gtgtggttgg	tgtgcttgtg	gtcccagcta	ctcaagaggc	tgaggcaaga	14820
gggttgcttg	agcccagaag	ttggaggctg	cogtgaactg	tgattgcacc	actgcacttc	14880
agcctgggtg	acatagcaag	accctgtctc	tgtggtggtg	gtgggtgggg	gtgggggaag	14940
ggattttaaga	aggttttgtg	aggtatgtat	tatttataaa	tgggctttta	actttaccct	15000
tcacatcttg	ggttgaaatt	aattgtatcc	attctcagtt	tttctgtctt	gctatatatt	15060
taaaacttga	gacttagagg	tcattggatgt	ctttctatga	aaagcaaatg	aagcagaggg	15120
ctgccttctc	ttgctgtaga	gggcacactt	gctgcagagc	atgttactgt	tttatgcatt	15180
gctaggcttt	gggagtgtg	acttgtatga	tcatagtact	tacaactatt	agttggcaat	15240
ttttaaactt	taactttaga	ttatatatgt	aaactcctgt	gttcctttgt	cactgataat	15300
ctgaacagaa	gccttgagata	aataattttg	aagtttttgt	ctgaacctct	gaaattttga	15360

ttgttatctc	atggttttgc	tgggaggaag	gagaaataac	aatggccact	tactgtgctt	15420
ctgtatgtgc	cagacagtat	gtgctagatg	tttcagaaac	gtgatttgta	atcctgacaa	15480
gaagccta	tgggtggtag	tgggtgctaa	ttgaacctta	tagatgagga	aattgaggct	15540
catggtggta	agtgaataac	ttgcaccaag	atcctatggc	tggatgagga	tagagcctca	15600
attcaagtac	gggtcttcca	ggtccaaacc	catgcaggct	ttgagaggta	aggaggtaga	15660
gaacgttgac	accccttctt	tgggtgtgtt	ttcagcaaat	acttgtatgc	atattaaaga	15720
ctgtctaccc	ttttgtcatc	ttgtgtcact	tgtgtcttcc	tttggtaacta	cccaaatttc	15780
tttcagcatt	tcagctttga	atttttattt	ttattttatt	taatttattt	atttttttga	15840
gatggagtct	cactctgttg	tccaggctgg	agtgcagtgg	cgtgatatca	gctcactgca	15900
acctctgcct	cacaggttca	agcaattctt	cctgcctcag	cctccttagt	agctgggact	15960
ggaggtgccc	accaccacgc	ccaactaatt	tttgtatttt	tagtagagat	agggttttac	16020
cttggttgcc	aggctgggtt	tgaactcttg	gcctcaagtg	atccaccac	ctcggcctcc	16080
caaaatgctg	ggattacagg	catgagccac	tgcacctggc	cagctttgaa	tttttagaat	16140
actgttctaa	acagaactat	attggaacct	ggaaaattaa	tctattgtct	ctaaatacca	16200
aagaaaaaca	tgaattttta	gtgggttgatt	atgggaacaa	ttttttttta	gatgggttca	16260
ctgaatggga	agcatttttt	ttttaattgc	ttgactattt	ctttaaat	ggagaaaaga	16320
ccattgcct	ctcagatttc	tggtaattgg	tcacattgat	catttatatt	gactgacagg	16380
ctgctttgtc	cacagctgaa	ggattgttta	atttttttta	aattataaga	gtaatatgtg	16440
ctcactgtaa	aattcacagt	acagaagcat	atgaactaac	taaaagtctt	tacctcttgt	16500
ctccagcaag	gagtaagtgt	ttcaacctga	agggttggtt	tgaattgtgt	tctgtggagc	16560
gtacttaaag	tgagtgaaga	agaaaaattt	atgtcaatca	tgatcattgc	agctgaagtt	16620
tttattgttt	caccccttaa	aggttattaa	aatagtatgt	agtttagtag	tcttgataat	16680
tttcccttaa	gatttattgg	ccagtatatt	aggattttgt	tttaaat	atatgtgagc	16740
ttagttttat	gctattttca	aataagacat	ttagaagaag	ataaaataac	attcctgtct	16800
tagtctgttt	tctgtctgta	taacagaata	gcacagactg	ggtaatttat	aaacagtaga	16860
agtttatttg	gcctgtgggt	ctggaggctg	ggaacttcaa	gagcatgggt	ctgccctttg	16920
tgtctgtgta	tcatatggtg	gaagggtggaa	aggcaagtgg	gtatgtcaag	acagagagca	16980
agaaggggct	tgaactcact	tttataacag	agtgactcca	gagatagcta	accacttttt	17040
gagagaatgc	attaatccat	tcatgagggc	agagcccttg	tgacctaatc	acctctcatt	17100
aggctctgca	tccttaaaact	ggtttttttt	tgtttttttt	ttttgagacg	gagtgtgcct	17160
ctgttgccca	ggccggactg	cggactgcag	tggcgcaatc	tcggctcact	gcaagctccg	17220
cctcccggt	tcacgccatt	ctcctgcctc	agcctcccga	gtagctggga	ctacaggcgc	17280
ccgccaccgt	gcccggctaa	ttttttgtat	tttttttagta	gagacggggt	ttcaccttgt	17340
tagccaggat	ggtctcgatc	tcctgacctc	atgatccacc	cgcctcggcc	tcccaaagtg	17400
ctgggattac	aggcgtgagc	caccgcgccc	ggccccctt	aaactgttgt	attggggatt	17460
aagtatctaa	cacaggaact	ttggaggata	catttaaaacc	ataagaattc	ctgtcatgca	17520
aatgaatcca	ttctagatga	aagagaatga	ttttagtttc	cattgaactt	tataaatagg	17580
ccttttctaa	ggtacttaca	gctgatatta	taaaatttat	atttgttttt	ataaaattgt	17640
atttgtattt	ctgtttgtac	aaatacaatt	atacactata	gttctctgct	gttagatttt	17700
ttttcttctt	tagcatgttt	ccaaagggtg	gaatgttgaa	agttgggtta	atgtcaatca	17760
gctttctttt	gtaaagtgtt	cattgacatg	tgaaccttgt	ctgagaatct	aaattttatt	17820
tcatgaaaga	agaaaacagt	atattctcat	ttaaaccaga	atttaacttc	atatacttgt	17880
ggctgtattg	ggagtatgcc	attgctgtct	gtttacaacc	tgacctactc	tacctactta	17940
gaagtaattt	gtgttatgat	agggtgtgctg	tgtgacata	tgctgaacat	atttgtaagg	18000
gtgttaagtc	attgaataaa	acgcttttct	cctcctttca	aataacattt	tttatttctg	18060
gttataaaaag	tcatacaagc	ttactgcagg	ttgttaaaaa	ggtataaaga	agaaaccgtc	18120
aatccattat	aatcctacag	tttagacttc	ctgtctcagc	ctctcagagt	gctgagatga	18180
gctagccatg	cccagcccct	caaaagattt	tttaaaaaac	aaaaatgagg	ttatacttta	18240
aaaaattcta	tattcctttc	acataacagt	gttatttttg	aggttttaga	atttccagta	18300
gcatttttaga	ttcagaaaca	agctgattca	tcctctactt	tgtacttttag	gcaagaaaag	18360
aattttacct	aaatagaatt	ttgaactgaa	aatctgtttt	tctaactttt	tatttaaaga	18420
atattgttcc	atgctttcac	agtagtgact	tttaattttt	atatttttta	ttttatttat	18480
ttagagatgg	gggtctcact	cttgttgctt	aggctagagt	gagtgcattg	gttctatttc	18540
tagctcactg	caaccttgaa	ctcctgggct	caagttaccc	tcctgcctca	gccttctaag	18600
tagctgggac	tacaggtgtg	caccactgca	ccaggctttt	tttaaggga	tagaaaatgg	18660
tagtgcttgc	atacaaaaat	ggcgtaggtg	catacatcag	cggacatcaa	gactatgttc	18720
agatcataaa	tgtacatata	tgtaccgatg	ccattttttg	acgcaaaaca	ataatggaaa	18780

ttgaactcta aactgaaatt tgaaacaagg gttctggggg gggccctctt gctgatttgt 18840
 aattgaatgt atagttcaat ttttcccat ctgttaagca aaagacaatt ctaatgttag 18900
 caaaaatcca catatccgtg cattgatcat ttttcccta attttcttta agagatgggg 18960
 cttctctcta tgttgccag gctggctgg aactcttggg ctcaaatgat cctccagcct 19020
 cagcctccca aagtgctgga attaataggc acaagctgct gtgcctggcc ctgtcatcag 19080
 tcatttaact tcatgcaaac tgagtagaat aaaactcgtc cttactgtac cttattgctt 19140
 ttgttttatt gttggaacct ccaatattgc gaaagtagac caaaagttga cttataggaa 19200
 aaactgatag caaaaaataat ttttctcttg ttgctgtatt tcatgcccac catccagttg 19260
 ttaaagccta ctgttaatth ctctcagcct cctcctttct gtccaggctt attctatgcc 19320
 attcttacct taactgtttt tagctttctc atagagtga ctttttaaat taaaataaaa 19380
 tatctgctcg tagctattata aaattcaagc agttcaacag aatttttcac taatagaaat 19440
 acttgtagct caaaagcagc tttattttac aaaccagcc caatttgtga ttagatttaa 19500
 cttgagaaaa catgaaatgt ctctcatatt gtttaaaaat atcataagtg gctgggcacg 19560
 gtggccttat cctataatcc caacactttg ggaggctgag gcagggtgat cacttgaggt 19620
 caggagtttg agaccagcca ggnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 19680
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 19740
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 19800
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 19860
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 19920
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnntt 19980
 accatgttgg ccaggctggg ctcaaaactcc tgacctcagg tgatccacct gcctgggcct 20040
 cccaaagtgc tgggattata ggcttgagcc tcgcctggcc tcctcataat tttttaacct 20100
 ttataaaaac cttttctaaa acccttttta ttttgaacta aatttagatt tactgaaatt 20160
 gtgaaatcaa tgtggagttc ttgtataccc tctttccgc ttttctaat agtaacatct 20220
 tacatacatg gtacatttgt ccaaatttag aaataaacat tggtagcagtg ttaactatag 20280
 acttaactcg gtttctctaa ttttttact aatgttctt ttctgttcta ggatctaatt 20340
 cagtatacca tattgtattt agttgtaggc catgttagcc accttcaatc tgtgacagtt 20400
 tctcagttct tccttctttt tcgttatctt gacaagtttg aagagtgtcg ataggtattt 20460
 tatagaatgt ccgtcagttg tctgtcagtt tgtatttgtc tgatgtattt tttttttttt 20520
 ttttgagatg gtgtctcgtc ctgtcgccta ggctggagtg caatggcatg atcttggctc 20580
 aatgcagcct ccacctccgg ggttcaagtg actgtcctgc ctacgtctcc caagtaactg 20640
 aaactacagg catgtgccac cagcctggc taattttttg tatttttagta gagaagcagt 20700
 ttcaccgtgt tgcccaggct ggtctcgtgc tcttgagctc aggcattcca cccgcatgg 20760
 cctcccaaaag cgtaggatt acaggtgtga gccaccatgc ctggccaata ttttgaggga 20820
 tatactttgg tgaggtcatg cagatattct gtttctcctt agttttatcg attaatatag 20880
 catttatcca gtaaatcttc cttgcagcaa ttattttttc tttttctttt ttccttaatt 20940
 ttttttttaa gagatgggat ctactctgtg tgcccaagtt ggaatgcagt agtgagttca 21000
 tagctcactg cagcctcaaa ctctgggct caagtgatcc tctgcctca gcctctcaag 21060
 tagctgggac tacaggcata gaccaccaca cccagctaatt gctggcctca tgtgatcctt 21180
 atgggggttt tgctatgttg ctgaggtgg tcttgaactt gctggcctca tgtgatcctt 21180
 ctacctcagc cttacaagta ggtgggaatt acaggtgtga gccaccacac ccagcattgc 21240
 agcaattatt aatgtagtgc tactggtcat tttctgtttt tctcatttct tcagcatgtg 21300
 ttattgactt gtctcttccc tccattttat aatcatttat actgctatga attcatgagt 21360
 atttattttg tgagtataaa tctaatacgt acttaattta ttttgtgcct caaattgttc 21420
 tggcttggcc attttttttt tttttttttg agacggtctc gctctgctgc ccaggctgga 21480
 gtgcagtagc gccatctctt ctactgcgaa cctccacctc ccgggttcaa gcgattctcc 21540
 tgccctcagc tctgagtag ctgggactac aggcgtgtgc cgccacaccc gtctaatttt 21600
 ttgtattttt agtagagaca gggtttcacc atgttagcca ggtggtctc gatctcctga 21660
 cctcgtgatc tgcccgctc agcctccaaa agtgctggga ttacaggtgt gagccaccaa 21720
 gcccgaccgg ctctgtatc cttttaacat gaggtgctgt catcattttt tccccctaat 21780
 attttggcca aaaatgttaa tcaaggatgg cacaaatttt ctgtagctgt atctcacaat 21840
 gaaagaggcc tgattaaaaa tgtaaaacta aaatgttctc tgatctctta gcacatgctt 21900
 tgtaaaaggc acagtgttag atccttgtat acgtagatga gtaagtgcac ttaccttcca 21960
 caccacaga tagctatgtc aaacgtaaag gtggagaaac acagacccca aacttctcga 22020
 gggtagaaaa tatgaggtta tagtagatta gaactacaaa aagctagagg aagttctgaa 22080
 ctggaacagc tggtataggat ttactagaat aatttacgag ggtgacaatt gtaaatcttc 22140
 ataggtttct tttttttctt ttctcttttt ttttttttga gatggagtct cgctctgttg 22200

cccaggctgg agtgcaatgg cgcagtcctct cctcactgca acctccgcct cctgggtcca 22260
 ggtgattctc ctgccttagc caccgaagta gctgggatta caggcatctg ccaccatgct 22320
 gagctaattt ttgtattttt ttttttagta gagacggggt ttcacatgt tggtcaggct 22380
 ggtcttgaac tcctgacctc aggtaatcca cccacctgg cctcccaaag tgctgggatt 22440
 acagggtgta gccaccgcgc ccagccaaat ttttatgggt ttctaaacta gcgtaattta 22500
 gtttttttca cttaagtcaa aattatatta ttgtaggata aaaacttagt gatccaaatt 22560
 catgaggaat gaagaataaa tacattttaa gtcttaccat ttgctaaatt agtcttggtc 22620
 ctttgtagca aaattctgtc cttgtgctct gtaattttat attgtatat ttctatcaa 22680
 catttttact gtgtgggtgt ttgtaaatta taaaaacgtt ttaaagcaaa ctcagaacaa 22740
 tgaattctca cgaatattca gtatatttac agttgagaaa taaactactt ctgtagtagg 22800
 taatttaaaa tgtcccaatg caagttaacg tgtcactgat cacgctattc aggtgtgtgt 22860
 ctttgataag gggagggtgg gaagtttggt ggtttgattt tatttgcctt tctcatgtga 22920
 ctgttgcctg gttagtaaac aaatgggttg cgagagaacc agtagtcttt tgcaaagatt 22980
 gtcttataca gagcactcaa ttcttcatat tatttataat ggctttaatt taagccttaa 23040
 attattagaa actcataaat aattttttta ttgtttttt tgagatggag ttctgcctt 23100
 attgtccagg ctgaagtaca atgatgtgat cttgactcac tgcaacctcc gcctctcggg 23160
 ttcaagtgat tctcctgcct ttgcctccca agtagctggg attacaggca tgcgctacca 23220
 tgcctggcta attttgatt tttagtaaaag acaggattgc accatgttgg ccaggctggt 23280
 ctggaactcc caacctcagg tgatccacct gcttcggcct cccagagtgc tgggattaca 23340
 ggctcactga gccactgtgc ccagccataa tgcgttaaaa taagagtgtt atatttgtaa 23400
 aacttaaaaa aatgtagtgg ttgaaaaagg taatttaaaa agaattgact attaatttct 23460
 tgaaaccata atgtaacttg tagtgcaatt aggaacctt catgtttctt tctttcttct 23520
 tttttttttt tttttgagat ggagttttgc tctgttggc taggctggag tgtgtgatgt 23580
 cagcgactg caacctctgc ctctgggtt caagcaattc tctgcctca gcctcccgag 23640
 tagctgggat tacaggcgcc tggcaccaca cccagctaat ttttgattt ttagtagagg 23700
 cggggtttca tcgtgttggc ctggctggtc tcgaactcct gacctcagg gatccactgc 23760
 acctggcccc cgttcatgtc ttttaaagct ttatggttgc tctgaaatag agttgttgat 23820
 tttttttttt tttttgagac tctcttttg cccgtgctgg agtgacgtgg tgtgatctga 23880
 gctcactgca acctccacct cctgagttca agcaattctc atgggtcagc ctctcaagta 23940
 gctgagatta aagctgcccc ccagctgcc tagtaattt tagtattttt agtagagatg 24000
 gggtttcacc gtattggcca ggggtggtctg gaacttctga cctcaggcat gagccactac 24060
 gcctagcctg ggttgttgat cttaagggtg atacttcagg caacatctga ggcccagtac 24120
 agtcctttac ttcaactggc tccagtacag caaattcagg gaatgtttt gagtggttac 24180
 tggatgcctg gcgtggagtt caggagatt ggtacattga gtccagttgt tgtgttgaaa 24240
 ctctgttta aaaacctccc tactaagtc cagctactca ggaggctgag gcctgagaat 24300
 cacttgaaac cctggaggca gaggttgagc tgaatcgaga tcgagccact gcaactcagc 24360
 ctgtgagaca agctgagact gtctaacaac aaaaaaaca cccccaataa aaccaacta 24420
 ctatggtagt atcaatgctg tgatagtctt cctttcttca tacaggtaaa ttcttaacat 24480
 atactcattg ttaatgttca gtgttcagta ttcttaagag tatttggggc caggcacggt 24540
 ggctcatgcc tgtactcca gcactttggg aggtgaggt gagcagatta cctgaggtta 24600
 ggagcttgag aacagcctcc aacatgatga aactcccgct ttactagaa atacaaaaat 24660
 tagctgggtg tgttagcaca tgtctgtaat cccagctact tcagaggctg aggcaggaga 24720
 attgcttgaa cctgggaggt ggaggctgca gtgacctgag attgcttcac tgcactccag 24780
 cctgggcaac agagcgagac tcttgctca aaaaaaaca aaaaaaaag aatatttggg 24840
 gccaggcatg gtggctcaca cctgtagtcc cagcactttg ggaggccaag gtgggtggat 24900
 cacttgagat caggagttgg agaccagccc gaccaacatg gctaaatccc gtctctacta 24960
 aaagtacaaa aattagcttg agcaacagag caagactctg tctcaaaaa agaaagaaga 25020
 atatttgggt taattaagaa ggaaccttat caatagtagt aaagtcagcc agctgaactg 25080
 ccaagtacaa attgttggtt ttaggtatca atcatttatt aaggataata ttctacaata 25140
 gcgatctttt taaaaattt aaaatctcaa actggaaagg atgtctagtt cattctatgc 25200
 ttcagtcctc tcttctgatt tactgttta gaagattttt gtttcttct ctgacttcta 25260
 ttttgctgct gactggcact tgggattttt aaaaaattat ttctctcata tataattaaa 25320
 gacaataagt ataacaataa gtataatatg gtaatttgct aaaacccaaa caatgtttta 25380
 agtaatgcat atcattatgt aaacctacgt aatagttgaa tattcacaaa gataatcgct 25440
 tatagaagtt ttatatcctc tcttctttgg cagtgcattt aaaacaaaaa aaataagttt 25500
 tatgtcttgt ttacatgtaa ataattttta tctaaattgt gacgtggttt tcaacttagc 25560
 atatttttga aagtaaatca aaaaggacaa aatacaaaat catgtatatc ttctacaaaa 25620

acgatataata aattctaagg tttttgtcct tttgaaattg cttaaaagaa tgcataagaac 25680
 tgggtgtctga gttgggaagg atctatgagg gatttccttg gagaccgtgg gtgaataata 25740
 atgttgtctt agttccatga aggaatctct ggggatagtt tttgagttag gcctggcaat 25800
 gttagagata cataaagaga gccttggttt atcactgggt gcggtggctc acacctgtaa 25860
 ttccagcact ttgggaggct gaggcgggca gatcatgagg tcaggagatc gagaccatcc 25920
 tggccaacac ggtgaaaccc gtgtctacta aaaatacaaa aattagctgg gcgtgggtggc 25980
 gcatgcctat aatcccagct actcgggagg ctgaggcagg agaactcact gaaccaggga 26040
 gttgaggtt gcagtgagcc gagatcgccg cactgcactc cagcctgggt gacagagcaa 26100
 gactccgtct caaaaaaaa aagcttggtt ttcaatggtt ctgaaaaatg ctttaataca 26160
 agtgtagagt gttagtcaag ttttgcactt ggataaacag cctgtgaatt tatcacattt 26220
 ctagtttata atatgggctt tcagaagtta tatgaacatt gttttgacgg gagaattcaa 26280
 gctggatgct agagaaggat cgtgagaacc ccttcatttg aggagtgcta tgaaattatt 26340
 tgatcttga attttttttt tttttttttt tttttttttt ttttttgagac agagtttctg 26400
 tcttattgcc caggctggag ctggaatgca gtggcacgat ctcggtcac tgcaacctct 26460
 gcctcctggg ttcaagcaat tcttctgcct cagcctacca ggtagctggg attacaggga 26520
 tgcccaacca tgcccagcta atttttgtat ttttaatgga gacggggtt caccatgttg 26580
 gtcaggtgg tcttgaactc ctgacctcaa gtgaactgcc tgcctcagcc tcccaaagt 26640
 ttgggattac aggtgtgagc cactgcgcct ggcctgatct tagaatttga aggagagact 26700
 aatatttcat gggcaaaaac aatgaaaagt tacctttctg tattctaata ctatagagga 26760
 gtgggattta tttagaatgt ttttaagtatc ttgggcagtc caagagtgcg tatcacttat 26820
 ttttcttttc cttctttctt ttttaagtga agttcactga tgttagagat cataggtggc 26880
 attgcctact ttttacataa ttttatcatg ttttagtgc tgtcagaagg gctgtggctg 26940
 tttgcagttt tggcctaaag catgcattgg ctttatagga gatgtagtct tcacagttag 27000
 ttgttatttg tagctgtgtt tttgtttttg tatagcttat agcaatgcag tgtgcttttt 27060
 attaacatca ttttcttttt ctttttgagc tgattattta ttcaagttac ttctgatttg 27120
 cgactcagg gttgaaaagt cttgccttct tcttaggttt gcagtaagtt gaaattgaaa 27180
 tgtctttaca attaatggtt caattaatgc tatgtatgtt ttctaggtag ataaaaattaa 27240
 acagttttat tcagaataag ttaattcttc cagaatttat atattttaaag actccaaata 27300
 tacatcccca gtggtatctt ggactgttaa atagaaaaat attgttgctc ttaaaagaaa 27360
 tttagtgaa gttgtgtata aagtcagaat gtcttaact tttggtcaga gtcaaacagc 27420
 agttccaata taggcagcaa gttaaagggt tagttgttg cctgtgttga aagcgacttg 27480
 atgaaaataa atcttttaaat taaactttag tagaataaaa agaaaaagca gagccagggtg 27540
 acgcagtgg tcatgcctgc agtctcagct actcagggtg ctgagggtgg aaggatcact 27600
 tgagtctagg agttttgaga ccaacctgga caacatagca tgactctgtc tctgaaaaaa 27660
 aaagttaata aaagaaaaag tagggtcttg gacaaacttc gttggccaat ggcatagttc 27720
 taaatgctga agctgacaga taaaggactt ttgacttaac agaactcaca gtgtccttca 27780
 ttagctttt caactacctt taaatttagc atgtttctg gccagggtgc gtggctcacg 27840
 cctgtaatcc cagcactttg ggaggccgag acgggcggat cacaaggcca agagattgag 27900
 accatcctgg ctaaacacggg gaaaccccgct ctctactaaa aatacaaaaa atcagctggg 27960
 tgtggtgcca cagcctgta gtcccagcta ctgggaggc tgaggcagga gaatcgcttg 28020
 aaccaggag gcggagggtt cagtgcctg agatggtgcc actgcactcc agcctggcaa 28080
 cagagcaaga ctgtctcaaa aaaaaagaa aaaaaataaa aaaacaaatt agcatgtttc 28140
 ccttctagag atcattgttt ctgagagcat ggaccaaaga ctctggggg ttaccaagac 28200
 cctctcaggt agcccatgag gtcaaaatat cctaataata ctaagatgtt agtattttga 28260
 aggaaatatt tacttggtta taatactaat ataaaagatg tttgcgtttt tcagtgtatga 28320
 cattggctct ggtacaaaag catgtgggtta aaattgtgc tggcttggtta cacatcaagg 28380
 cagcgctaag ctccaaattg tactcatggt gatggcattc tttacctctg tgccctcaca 28440
 ggaacaaaa caagccgtgc catttttatt gaagattgtc cttgacaaaa cagttaaaat 28500
 gattaatttt tgaaaaatgt tgatccatga gtattccttt aaaaatattt gtgaagaaat 28560
 gggaaagttc cataaaacaa tgtttttttt ttgttttttt tttttttttt ttttgagaca 28620
 gattctggct gtgttgccaa ggctagagtg cagtggcgct tggctcccag gctcaagctg 28680
 ttctccact tcagcctccc aagtggctgg gacctcccaa gtggatgcgc catcatgcct 28740
 ggctgatttt tgtatttttt tgtagtgaac aggtctcact gtgttgaca ggctggctct 28800
 aaacttctga gctcaagcga tgcattgtgc tcagcctccc aaagtgtctg agaaagcact 28860
 ttttactgca tactggctag tgtgttggtt attttgaga aaagaaaagc atttgtagtt 28920
 ttttgagttg taagctgagc taactgcttt attttttct gtggaacacc atttcttttt 28980
 ttttttttga gatggaatat tgctttgttg cccaggctgg agtgacagtg cacaatctcg 29040

66

ttcttaggtg	gtgctttgtg	gttcatatga	tgatagcctg	gtctgttcat	tcatatatct	32520
tttcacgaga	aatattttta	ttccattctg	aataaaat	catggcaggt	acttgcaaga	32580
agcagttata	attttaaa	ttaacattag	gttaaaaaat	tgacaggaaa	catatattca	32640
caggtaaaac	ttgtacacaa	atgttcatgg	cagcattatt	cataatagcc	aagaagtgg	32700
aacaacccaa	atcaatttat	gaatggataa	aatgttgtat	atgtgtagta	catgtaatat	32760
tattcagcca	ataaaatggg	ccaggcatgg	tggctcacac	ctgtaatccc	agcactttga	32820
gaggctcagg	cagggggatc	actagaggtc	aggagtttga	gaccagcctg	accatcatca	32880
cgaaacccctg	tctctactaa	acgtacaaaa	attaggcagg	cgtgggtgatg	cacgcctgta	32940
gtccctacta	ctcaggtggc	tgagtcatga	ggattgcttg	gaccccgga	gacagaggtt	33000
gcagtgcagt	gagatcatga	cactgcactc	cagcatgggc	aacagagcaa	catcctgcct	33060
caaaaaaaaa	aaaaaaaaaa	aaaagaagta	ctgttacatg	gtacaacatg	gatgaacctt	33120
gaaaacattc	tgctaaatga	aggaagacag	acacagaggg	ccacatat	tatgattcca	33180
tttatacgaa	atgtccaaaa	ttggcaaatc	taaagagaaa	gtagattagt	ggttgccagg	33240
gagtgaagac	gggttctttc	tggagtgaag	aaaatgtcct	ggaattcgtg	gttgtagttt	33300
gcaaatctgt	gaatgtataa	ggaccactga	attgtccact	tcaaaagggt	gacttttatg	33360
ttatgtgcat	tatatctaaa	aaaaaaaaatca	taattaggaa	gcaagattga	cttctaagaa	33420
aaagcggagt	gaaattgttg	ttttgtgggtg	aataaattgg	gtgggtgggt	cgcaagagtt	33480
ttgctgatta	gtgattagaa	aaattattca	taatcattga	aaatataaaa	tatttttcta	33540
tatgatgtat	gtaaagaatt	tggcaagaga	tgatgttttg	aaaaataaaa	gaatggtat	33600
tgtagagatc	ttaaggaaag	aaactacagt	taagtagtgc	tttghtaatca	gaatataag	33660
taagtactga	aagtggatgg	agtggctgtt	gtcagcatgt	tatactttat	acatttcatt	33720
cataaaattg	gactgtagat	aaaagtaaac	ttttttttta	tttactcttg	aacaacagtt	33780
tttttttttc	cacttagact	tgcactgtct	cactgaaca	atacatttaa	ttgttaatta	33840
tttccccctt	caggatgata	catatacaga	aagctacatc	agcacaattg	gtgtggattt	33900
caaaataaga	actatagagt	tagacgggaa	aacaatcaag	cttcaaatag	taagtgactt	33960
ggctagtaat	ttttttgaaa	tttattttgg	taaatttgta	atgtattgtt	attttgtata	34020
tattttactat	gctaacaaaa	ttgaatgtaa	aatgtcttaa	gattcatgta	cttaagatag	34080
aatggtagaa	taagaattac	ttagattaaa	aataatattt	tcaagattac	ttaagcctca	34140
ttgaattttc	tgttcatgaa	gcagagaaac	tcattgtttta	agtcaaaact	ggctctcatc	34200
tttttctttt	atcagtggaa	atctaagttc	aagtttacct	tgtcctacac	tgcaaatgct	34260
atagaccatt	tttgtttgtc	ttttactgtg	ctaagtgcac	ggaacattaa	aggaacccta	34320
ggaagagatt	cttcatatgt	ggctcagttg	aagagaagta	cttatgtagt	tctaagtatt	34380
tttatttagat	agtgtgcacc	aactctgtag	aaacacagaa	ttttgttggg	aaaaggaaact	34440
tagtttttgt	aacatgttca	ttttactgct	caaaaaaacg	aatgctgaaa	gatttaataga	34500
cttgcctaca	gttactggta	gaaccaagtg	accgaagctc	tgtcttcaat	attttgtgtc	34560
tgtgtgccat	cctatcccc	ttatccatct	ttacaccccc	agcccccact	taaataatagg	34620
caattataat	agttcagttg	tgcctcttca	gtatgggtct	gagtcctgtc	agtgtgggca	34680
tatctgtggt	cttttaaaaa	ataaatctct	cagtattttt	cagagtaggc	tattagcaag	34740
aagtaggcta	taaacacagg	aaaccagtga	ctgccccctt	tcattggaact	gatgacacat	34800
ggaattggaa	ggagtccctg	attaggagtc	agaagactta	gatttgttgt	cttgggtcta	34860
gtattttacct	gttagagaat	catgggtttg	tgtctctggg	gaaaaggccg	aagtaacctt	34920
gagaccagct	ttcctttcta	aaatgtgtgt	gatgacacct	gattttactaa	tttataagct	34980
agttgtgaga	accaactgta	atagctttgt	gtatgtgaca	atagctgtga	aagccctttg	35040
taaaacttttg	ggcagcatat	agatactact	tatgatatga	catgccccaga	taaatgggtg	35100
tttgataggt	taagtgtgtc	ctttttctta	catgactctg	atgaggaaaa	gaaggtatgt	35160
taacaaaaga	taggtggctg	tggatattga	tataagtaaa	cacacttgat	gtgtcaaat	35220
aggacttgca	aggatttagt	tttcagaaat	agcttgaaat	actttcaatc	agtgaacaaa	35280
ttaccctcca	tattttttcc	cacgatataa	gtacagtctc	aaccttttat	ttggcaccat	35340
aaagagcaca	taaagatcta	cccaaaactg	tacttttaaag	cactggtagt	gaataattgt	35400
attatgtgtg	atcattgggtg	tttataagat	ttgggtgtgt	attcgtgtgt	gaaacattca	35460
tatttttgta	ctttcctgtg	gctggaagg	atcttatag	acactgtctt	tcacttttgt	35520
ctgtctttca	tcttttaata	gaatttcttt	tccatgcctg	aaggccctcat	tttgaaacatt	35580
ttgtttgttt	gtttttttat	tttttgagat	acagtattgc	tctgtctccc	aggctggagt	35640
gcagtggcgc	gatttgagct	cactgcaacc	tccgcctcct	gggttcaagt	gatttctcctg	35700
cctcagcctc	cctaataagct	gggattacat	gtgtgtacca	ccatgcccg	acaatttttt	35760
tttttttgag	atggagcctt	gctttgtcgc	ccaggctgga	gtgccagtgg	tgcaatcttg	35820
gctcgctgca	gcctccgcct	cccaggttca	agcagttctc	ttgcctcagc	ctcctgagta	35880

gctgggatta caggcgtgcg ccaccacacc ctgctaattt tttgtatttt tagtagagac 35940
 agagtttcac catgttggtt aggctggctt cgaactcctg acctcgtgat ctgcctgact 36000
 cggcttccca aagtgcctggg attacaggca tgagccactg tgcccagcct tccgataatt 36060
 tttgtatttt tcgtagagat gggatttcgc catgttggtt aggctggctt caaactcctt 36120
 acctcaagt atccacccgt cttggcctcc caaagtgcgt ggattacagg cgtgagccac 36180
 cacgcctggg tttttgaaca tttttaagaa gcttaccatt ttttcgaaat agctagtacc 36240
 attttacaca taacttcagc taggcatgtt gcctcatgcc tgtaatccca gcactttggg 36300
 aggccgagggt cagagagtca cttgaggcca ggagtcaaca tagctcctgt gaccagcctg 36360
 gtcaacatag agactctatc tctaccaaaa aaaaaaaataac cagggtgtggt 36420
 ggtccatgcc tgtagtccca gctcccaggt agactgaggt gggagggaatg tttgagccca 36480
 ggacttcaag gctgcagtga ggcaagattg caccattgca cccagcctt ggggacagag 36540
 tgagagaccc tgtctcaaaa acaaaataag gctgggcgca gtggctgtcc gggcgtcgtg 36600
 gttcacgctt atagtccatg cactttggga ggccaagggt ggcagattgc ctgagctcag 36660
 gaggtctaag accagcctga gcaacatggc gaaacctcat ctttgcaaaa catacagaaa 36720
 aaaacaaaa aaaccacaaa acctctagtt gccagttatt ttttttattt attcctagt 36780
 attcttcttt ttttcttttt tctgagacaa aaatttcact ttgtctccct cgctagagt 36840
 cagcggctcag ctactacat gattctttta gagacatgtt aattctttat attgagctga 36900
 agcctgtttc ttttacttct gtctcttctt attctctcgc cttgtagagc tgcctgaatc 36960
 agattaattc ctcttttatt ggcaagcctg cccttcagat tgatcttatc acaacctttc 37020
 ttctacctct gaagtccca tttcttctg taatgatatt ttcagaacct tgtgcaattt 37080
 gggttattct tacattttat aaatgccttt tattaattt gatttcttaa atcaagtatg 37140
 agatataaca catgaggtaa atctgtctt gatttggagc ctgaatgaat ttctctcttg 37200
 aacttcaagg gctcatggcc ctttcttatt attaatcaaa gacaaccatt tgttgtttca 37260
 gtagctatat tatttctagt ttgggtctta aggtttttga tttgcttgtt ttttcttttt 37320
 tctttttttt ttttttgaga cggagtcttc ctctgttgcc ccagactggg agtgcaatgg 37380
 cgtgatctcg gctcactgca acctccgcct cccaggttca agcgattctt ctgcctcagc 37440
 ctccctagta gcagggatta caggcatgtg ccaccacgcc gggctaattt tgtattttta 37500
 gtagagatgg ggtttctcca tgttggtcac gctggctcgc aactcccagc ctgaggtgat 37560
 ccgcctgcct tggcctccca aagtgcctgg attacagtcg tgagccacgg cgctggccg 37620
 atttgcttgt ttttaattaa aataggggcc ttggccaggt gcagttgttc accctgttaa 37680
 tcccagtact ttgggaggct gaggcaggca gatctctga gttcaggagt tcaagaccag 37740
 tatgggcaac atggtgaaac cctgtctcta ccaaaaacac aaaattcagc caggcatgg 37800
 ggtgtgtccc tgtagttcaa ggtactcagg aggtgaggt gggaggattg cttgagccc 37860
 gagatggagg ttgcggtgag ccaagattgt gccatttgca ctctagcctg ggcaacagag 37920
 cgagaccttg tttcaaaaaa aaaaaagaag aggttctcac tttacacttc tgtgactgg 37980
 gttttaaaaa tctaaacaca ggccgggcac ggtggctcac gcctgtaatc ccagcacttt 38040
 gggaggcaga ggcacgcaga tcacaaggtc aggagttcgt gaccagcctg gccagcatgg 38100
 tgaagcccat ctctactaaa aatacaaaaa aattagctgg gcatgggtgg aggtgcctgt 38160
 aatcccagct acttgggagg ctgagacagg gaaatcactt gaacccagga ggccggagatt 38220
 gcagtgaagc aagattgcgc cattgcactc cagcctgggt acagagcgag actccgtctg 38280
 aaaaaaaaaa aaaaaaatct aaacacaaga ttttactttt aatcctatca tttcctcttg 38340
 cttggcttca gtaatccttc aagttttcta ggtcttttca aaatcttgat tctgttgatt 38400
 tataatttaa ttatcttttc ctttcagctt ttctgttca ggtgtgacat ctgggtcttt 38460
 atctgagttt tattagatta taaaacattc agcaagatag ggcagggtact ggtccagtt 38520
 gtacaccatg gaaggcctct ttctgtgatt gttcattcat gaggctttat gaaaatgtct 38580
 acattacacc aggcacttgg aggttacaga gatgaataaa acatagtcca ttaggaggca 38640
 gacaatggga gagacaaaca tgggaaaaag ttactctgat tatgaggagt aatgagaatt 38700
 acatatgaag gaaagtattg ttagtactgt taggatttag tgtcaggaaa gttttcagag 38760
 tagcaaggaa acatcagaaa ttttactctt tctgccaggc atggtgcatg tattattctg 38820
 ttctcacact gccacaagga actgaccaa actgggtgat ttattaaaaa aaagggttaa 38880
 ttgactcata gttctgcatg gctgaggagg cctcaggaaa cttactgtgg cagaaaggga 38940
 agcaggcacg tcttacatgg caggaggcga gagagtgtga aggaagtga gggggaagag 39000
 ccccttatga gaccatcaga tctgtgaga attcattcac tatcactcga atgggggaaa 39060
 ccgtcgtcat aatccaatca cttctccata atccaatcac ttccctcagt gattacaact 39120
 tgagatgaga tttgggtggg gacacagagc caaaccatat cagtgcctgt agtccagtt 39180
 acttgagggc tgaggcagga ggaacacttg agccaggag ttcaagatct gcctgggcaa 39240
 catagcaata cctccatttt ggataaaaag gaaattttac tttttgggtg ccattgctta 39300

```

gtttaatcag ctgtaacttc ttgttgactt ttagtcaaaa aacaattttt ccttctatct 39360
ttgtgaaaga ggttggtgag caaggaagaa aaggaaactt gctttattga gcagcttcta 39420
tagtcaggca catTTTtaca acattagttc atttaaacc ctttagctgt tgtacaagg 39480
gaatgctatc tagcatttac agatgaagaa actgttaggt gactctccct aatattaaat 39540
aaccaggaac ctggatttga tgttttgaag tcagggtagc ttgatcctcg agttcatgct 39600
tcctccaagg atacactgaa agactttgag cctctttttt tttttttctc tttttttgag 39660
acaggatctg gctctcttgc ccagagtgcg gtggtgtgat ctgagctcac tgcaacctct 39720
gcctcctggg ctcaagcgat tctgcctcag cctctcgagt agctgggacc acaggcgac 39780
gccagcatac ttggctaatt tttggatttt tagtagagac agggtttcac catgttggtc 39840
aggctggtct cgaactcctg agctcgtaat ccgcccgtct cggccccaca aagtgtggtg 39900
attacaggcg tgagccaccg acccagtcct aacagttttt taaaaccag aactataatg 39960
caataatgtt agcattttgt ttgggagttt gagcctaaat ggttgaagtg cagttaaattg 40020
ttcttaaaat acgttttatg aaagtatttg gagtctcttc cttacatttt tttctctagc 40080
atgaagacaa cacctagcca ggcatggtgg ctcatgccag taatgccagc actttgggag 40140
aatgcttagg gataattgct tgagtcagg cagcctgggc aatgtagcga 40200
gactctgtct ctacaaaaaa gaaaaaatta gccgggtgtg gtggcatgtg cctgtatgcc 40260
cagctactca ggaggctcag gtggaaggat tgcttgaggt gggaggttga ggctgcagcg 40320
agccatgatc atgccactgt actcagcctg gatgacagaa tgagacgctg cttgagaggg 40380
gaaaaaaaag acacctgctt gggatgatta aagttctgtc ttgactggta gttatttgaa 40440
ttaggtccct ccagtgtctt taatcatggt agaatgtgct agcaagtga tttgtcttac 40500
atggaagagt tctgtgttca agggctttcg gccagtggca ttcctaaaca cagtgttaaa 40560
ggcgttaggg aatgtgaaaa gtatgacata gttcctgtct tcaacagctt gtaattttag 40620
tattattatc gtaagctcaa ttgtaggtac tacttctttt ctggactttc aggtgcttat 40680
taccgtgcaa tttagtggtg tgagttgagg actaatgttt ctatatcaca tcctgataat 40740
ctccacagtt atgaaaacta aactatttcc cctccctcct acacttttcc ccaactttat 40800
tttaattggaa ttgtttggat ttcttgattg ttttgtaata gtgggacaca gcaggccagg 40860
aaagatttcg aacaatcacc tccagttatt acagaggagc ccatggcatc atagtttgtg 40920
atgatgtgac agatcaggta agttccaaga ggagatttg t tacagtga caagtaggaa 40980
gccattattt gattaatgtc agattcattt actactcat atataagcca tcagatttaa 41040
ttttatggca gaaaactttg tccactctca aatataaatg tgaatcactt aaaagacatt 41100
tgttttctcg taataaataa aagattagta attagtttta cgtttgtctt caagggattc 41160
tggttgattt tattgtcaac taaataactt tgatcaaata gccaaagact taacatatag 41220
gcaagagttt gtagggaatc gtgagttgct tggcttatac tgtgttcttg gtgttaagta 41280
ttaacaggaa tatggcctgg taattagaac ttgtccatca gaattgccaa aagtgggatt 41340
cgggggtctc tgccatgga ggatgtggtt cagaaataaa gaatttgaat aggataagct 41400
gtaggaggat cttagtatga gaatgagat ctgaagatta gctgtgagag agggcagagc 41460
gatggaggga acaatgtggg acagtgtgaa aatataatgc caggggccat aacttttttt 41520
gttactattt ttttaaatca gaaacttaga tttcagtgtc ctttctatca aagaaaagga 41580
caaaagataa acgttcaaaa ttggaattta tttttctttt ggcaaattgt aaatctcacc 41640
tctaattgaga aatcatagct aattaggaga taacttacat gtaagcattt agattcagtg 41700
ccattagaag tgctgggtgg gtgatattct caggagaaaa aaatgatgct agtttaaaaa 41760
atctctacta ttaccgtgaa atatttttaa atgaaaactt tcgtcctcta aatatgactg 41820
tggaagagaa aatgagtata ttttaataaca tcttttgaca tctctagtag taacagtagg 41880
tcatcttatt cataaacc aaattttacca aatttcaggc caggcgaggt ggctcatgcc 41940
tgtaatccca gaactttggg aggccgaggt gggcggtatc cctgaggtca ggagttagag 42000
actagcctcg ccaacatggc aaaatcccat ctctagttaa aatacaaaaa tttagccaggc 42060
gtggggggcc gtgcctgtaa tctagccac ttgggaggct gagacaggag aatcgcttga 42120
acccagcggg cagagggtgc agtgagccga gatcgcgcca ttgcaactca gcctggatga 42180
cagaacaaga ctttgtctca aaaaaaaaaa aaaaaaaaaa aaaaaatta atcaaatttc 42240
aaaaccaggt tttgtagtac atttaaattg catattccaa agcagttggg tttgcctgct 42300
ttgcagttta atattaagct atacttcctt ttcaataaag gtattttcat cgttaagcct 42360
gtaaattcta gtttgtcatt gtttagatat ttatagtcatt ttaatatata ctgtttacg 42420
ccagctgcaa tggctaacac ctgtaaactc agcacttttt gagggccaagg tgggccgatt 42480
gagctcagga gttcgagacc agcctgggca acatagttaa actccatcta taaaaaaat 42540
ccaaaaaaa aaagacaggt gtggtggcat gtgcctgtag tcccagctat cccggaggcg 42600
gaggcgggag gatggcttga gcttgggagg tcgaggtgtc agtgagctgt gattgtgcca 42660
ctgactccg gcctaggtga cagagcaaga ccctgtctca aaaaaaaaaa tctcttcact 42720

```

ccttagcagtg ggttatattg tagctagagt tgtctcacta gctctttgtt atttgtctgt 42780
 taggtcagga acgatgtttc tgtttattcc agaactatat tatcgaacta tattatcagt 42840
 ctttcaaagt tcttttttag agtccttcaa taatgttaaa cagtggctgc aggaaataga 42900
 tcgttatgcc agtgaaaaatg tcaacaaatt gttggtaggg aacaaatgtg atctgaccac 42960
 aaagaaagta gtagactaca caacagcgaa ggtatgttta aagtttaatt ttcatactga 43020
 atttgaagggt gttgaattat gtatgggttc tgcagtaaca gtaaggccac agccttttaa 43080
 aaatagtgc actagaatac tgtgacagtg acaatttgtg tagcatctgt ttggatccaa 43140
 tgaacttagt tcttcacgct ccattatgga tggtagaaat gcagtaagaa ttagtgaaaa 43200
 agatttttca gtgttaattg tgctcatta ttctcttagg aatttgctga ttccttgga 43260
 attccgtttt tggaaaccag tgctaagaat gcaacgaatg tagaacagtc tttcatgacg 43320
 atggcagctg agattaaaaa gcgaatgggt cccggagcaa cagctgggtg tgctgagaag 43380
 tocaatgtta aaattcagag cactccagtc aagcagtcag gtggagggtg ctgctaaaaat 43440
 ttgcctccat ctttttctca cagcaatgaa tttgcaatct gaacccaagt gaaaaaacia 43500
 aattgcctga attgtactgt atgtagctgc actacaacag attcttaccg tctccacaaa 43560
 ggtcagagat tgtaaatggt caatactgac ttttttttta ttccttgac tcaagacagc 43620
 taacttcatt ttcagaactg ttttaaacct ttgtgtgctg tttcccgtgg ttggttagaa 43680
 aatccttggt gctttcctga taccagactg tttcccgtgg ttggttagaa tataatttgt 43740
 tttgatgttt atattggcat gtttagatgt cagggttagt cttctgaaga tgaagttcag 43800
 ccattttgtg tcaaacagca caagcagtgct ctgtcacttt ccatgcataa agtttagtga 43860
 gatgttatat gtaagatctg atttgctagt tcttccttgt agagttataa atggaaagat 43920
 tacactatct gattaatagt ttcttcatac tctgcatata atttgggct gcagaatatt 43980
 gtaatttggt gcacactatg taacaaaaca actgaagata tgtttaataa atattgtact 44040
 tattggaagt aatatcaaac tgtatgggtg taagtattgt tttgattctt atgggttaag 44100
 ggaaatagag ccttgcatga tattcaacac agccatttgt gtgtgcacaa tgcaaaactaa 44160
 ggtattctag acctatctta gagcagcatc cagtatttgc tttctagata atatgcccaa 44220
 taacatgacc tagaggggct tctgtgctgt gtagggtatt aaccaacttc agtgggtcag 44280
 ggagctcaaa ctatatgtaa aacaagttaa gaatgtatgc tatctagccc gttatctctg 44340
 atccttctct aaaaccattt gaaatagctt cattgatcaa catttcataa atgcatctgt 44400
 ggtagaggta gaaagcagca cctttcctaa ttggcaaatg atcagactaa tgtgtgctaa 44460
 tgtttttctt ccattgcttc agtcagattc aactatttta tctccacag ttgcttaact 44520
 tgggtgttga ggagggttta agcattaaga taggaagcag gaaatttgat tgccttaaat 44580
 tttagaaatta tatccctaaa aattaaaaca tgaatactgg gtggtaatga taattgaggc 44640
 aaatgtattt attttggtga cattttgcat atatgaagat tttctgaaat aggaccttca 44700
 agatcctagg gggttttgtt tggtttttaa ttgtgaggaa taaaaaatct tctgccaca 44760
 ctggcatttt aagggtactg aggtcaaacg ttgtttcctt aggttgaaat agcagccaaa 44820
 acattcttca cgcaggggct tgggatatgg ctgctggcaa cacattttgt tgtgggctcc 44880
 ttaatttaat gataaaattt aagctaaaca caagccaaaa atgaataggt ttttttaatt 44940
 tttatttttc actaaacagg caattgaaat acatgggtaca aaaataagtg gtaagataat 45000
 tgtaaaatga aatggacaga atattcaatt ttccatctat gaaaatttca caataaaaat 45060
 catagtttac tttgtattat aggcgtgctt ggtggatcta ttcatcctca cataaggcaa 45120
 ctgacaaaatt cctgaagtta ccaatagtta ttttggtgaa gatctttaat gcttcagaag 45180
 ttttgttttt gccttaatac agtataaagg gggaaagagt tcagaaacta ttttctaaag 45240
 tagctaaatg acacaaaaca aatgtcaaga tactgtgatg ccatgccgtg cacttcattt 45300
 ttacacagta aaagtgtttt aaattgtcag cttattcttg gtgagttagc ggaacatta 45360
 catgaactta agatgagcat atttacagac ttaagtttgg aaaattccag cgttcttttc 45420
 cccatggcag taaagattgg gatttacaac aaatttcagc atgccttaag atttgcttct 45480
 atgtatacgc caataaatgt ggttctggaa aaaatatata cccctttata ccccatctt 45540
 caagtacaaa cggttcaaag ctactacagg ttttaataat ctgttcactt agtaaaagga 45600
 attaccactt gttctaaata taagggtgct ccaataaatta gtttacatag tgaagaagag 45660
 tgttctttaa tctaagcagc tgcacactct gtgaaatcct ttcagaatga tagtcattgt 45720
 ggtctgagca gtaatttcct attcttcgac cttggattga atttccctta gcctacatct 45780
 tgcccttcca gcatacttta cctcaaacct tctttgtgtt ccattcccac ctaagcttca 45840
 aaatagccct gtgttgacgt cgtcttccat ttgctgagct tacctatgga tctccaagaa 45900
 cccagatctt gaaactgctg atccagcttt gagtatcatc acttccctgt ggatttaact 45960
 tccatttaatt ttaagggaact actaagttat tccagtgtgg catcacagtg cagtttagca 46020
 gctcagctac ttgactctaa tttggccatg 46050

<210> 4
 <211> 222
 <212> PRT
 <213> Human

<400> 4
 Gly Gly Cys Gly Ser Lys Gly Gly Gly Gly Gly Gly Ser Cys Ser
 1 5 10 15
 Asp Met Ser Ser Met Asn Pro Glu Tyr Asp Tyr Leu Phe Lys Leu Leu
 20 25 30
 Leu Ile Gly Asp Ser Gly Val Gly Lys Ser Cys Leu Leu Leu Arg Phe
 35 40 45
 Ala Asp Asp Thr Tyr Thr Glu Ser Tyr Ile Ser Thr Ile Gly Val Asp
 50 55 60
 Phe Lys Ile Arg Thr Ile Glu Leu Asp Gly Lys Thr Ile Lys Leu Gln
 65 70 75 80
 Ile Trp Asp Thr Ala Gly Gln Glu Arg Phe Arg Thr Ile Thr Ser Ser
 85 90 95
 Tyr Tyr Arg Gly Ala His Gly Ile Ile Val Val Tyr Asp Val Thr Asp
 100 105 110
 Gln Glu Ser Phe Asn Asn Val Lys Gln Trp Leu Gln Glu Ile Asp Arg
 115 120 125
 Tyr Ala Ser Glu Asn Val Asn Lys Leu Leu Val Gly Asn Lys Cys Asp
 130 135 140
 Leu Thr Thr Lys Lys Val Val Asp Tyr Thr Thr Ala Lys Glu Phe Ala
 145 150 155 160
 Asp Ser Leu Gly Ile Pro Phe Leu Glu Thr Ser Ala Lys Asn Ala Thr
 165 170 175
 Asn Val Glu Gln Ser Phe Met Thr Met Ala Ala Glu Ile Lys Lys Arg
 180 185 190
 Met Gly Pro Gly Ala Thr Ala Gly Gly Ala Glu Lys Ser Asn Val Lys
 195 200 205
 Ile Gln Ser Thr Pro Val Lys Gln Ser Gly Gly Gly Cys Cys
 210 215 220